

## Description

# BIOINFORMATICALLY DETECTABLE GROUP OF NOVEL REGULATORY OLIGONUCLEOTIDES AND USES THEREOF

### CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/707975 filed 29-Jan-04, U.S. Patent Application Serial No. 10/707147 filed 24-Nov-03, U.S. Patent Application Serial No. 10/604985 filed 29-Aug-03, U.S. Patent Application Serial No. 10/651227 filed 29-Aug-03, U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, and U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03. This application also claims priority from International Application Number: PCT/IL

03/00970, filed 16-Nov-03, the disclosure of which application is hereby incorporated herein by reference. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; This application also is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/708953, filed 2-Apr-04, and U.S. Patent Application Serial No. 10/707980 filed 29-Jan-04. Both of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Oligonucleotides and Uses Thereof"; This application also is a continuation in part of and claims priority from U.S. Patent Application Serial No. 10/708204 filed 16-Feb-04, entitled "Bioinformatically Detectable Group of Novel Regulatory Oligonucleotides Associated with Alzheimers Disease and Uses Thereof"; This application also is a continuation in part of and claims priority from U.S. Provisional Patent Application Serial No. 60/521,433 filed 26-Apr-04, entitled "A Microarray for the Detection of MicroRNA Oligonucleotides"; U.S Patent Application Serial No. 10/708953, filed 2-Apr-04, entitled "Bioinformatically

Detectable Group of Novel Regulatory Oligonucleotides and Uses Thereof is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/707975 filed 29-Jan-04, U.S. Patent Application Serial No. 10/707147 filed 24-Nov-03, U.S. Patent Application Serial No. 10/604985 filed 29-Aug-03, U.S. Patent Application Serial No. 10/651227 filed 29-Aug-03, U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, and U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03. This application also claims priority from International Application Number: PCT/IL 03/00970, filed 16-Nov-03, the disclosure of which application is hereby incorporated herein by reference. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; This application also is a continuation in part of and claims priority from U.S. Patent Application Serial No. 10/707980 filed 29-Jan-04, entitled "Bioinformatically Detectable Group of Novel Regulatory Oligonu-

cleotides and Uses Thereof"; This application also is a continuation in part of and claims priority from U.S. Patent Application Serial No. 10/708204 filed 16-Feb-04, entitled "Bioinformatically Detectable Group of Novel Regulatory Oligonucleotides Associated with Alzheimers Disease and Uses Thereof"; U.S Patent Application Serial No. 10/708204, filed 16-Feb-04, entitled "Bioinformatically Detectable Group of Novel Regulatory Oligonucleotides Associated with Alzheimers Disease and Uses Thereof" is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/707975 filed 29-Jan-04, U.S. Patent Application Serial No. 10/707147 filed 24-Nov-03, U.S. Patent Application Serial No. 10/604985 filed 29-Aug-03, U.S. Patent Application Serial No. 10/651227 filed 29-Aug-03, U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, and U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03. This application also

claims priority from International Application Number: PCT/IL 03/00970, filed 16-Nov-03, the disclosure of which application is hereby incorporated herein by reference. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; This application also is a continuation in part of and claims priority from U.S. Patent Application Serial No. 10/707980 filed 29-Jan-04, entitled "Bioinformatically Detectable Group of Novel Regulatory Oligonucleotides and Uses Thereof"; U.S. Patent Application Serial No. 10/707980, filed 29-Jan-04, entitled "Bioinformatically Detectable Group of Novel Regulatory Oligonucleotides and Uses Thereof" is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/707975 filed 29-Jan-04, U.S. Patent Application Serial No. 10/707147 filed 24-Nov-03, U.S. Patent Application Serial No. 10/604985 filed 29-Aug-03, U.S. Patent Application Serial No. 10/651227 filed 29-Aug-03, U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Applica-

tion Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, and U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03. This application also claims priority from International Application Number: PCT/IL 03/00970, filed 16-Nov-03, the disclosure of which application is hereby incorporated herein by reference. All of the afore-said patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S. Patent Application Serial No. 10/707975, filed 29-Jan-04, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/707147 filed 24-Nov-03, U.S. Patent Application Serial No. 10/604985 filed 29-Aug-03, U.S. Patent Application Serial No. 10/651227 filed 29-Aug-03, U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03,

and U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03. This application also claims priority from International Application Number: PCT/IL 03/00970, filed 16-Nov-03, the disclosure of which application is hereby incorporated herein by reference. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S. Patent Application Serial No. 10/707147, filed 24-Nov-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/604985 filed 29-Aug-03, U.S. Patent Application Serial No. 10/651227 filed 29-Aug-03, U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03, and U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02. This application also claims

priority from International Application Number: PCT/IL 03/00970, filed 16-Nov-03, the disclosure of which application is hereby incorporated herein by reference. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; International Application Number: PCT/IL 03/00970, filed 16-Nov-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/604985 filed 29-Aug-03, U.S. Patent Application Serial No. 10/651227 filed 29-Aug-03, U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03, U.S. Patent Application Serial No. 10/345201 filed 16-Jan-03, and U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and



Uses Thereof"; U.S. Patent Application Serial No. 10/604985, filed 29-Aug-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation of U.S. Provisional Patent Application Serial No. 60/468251, filed 07-May-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof", the disclosure of which is hereby incorporated herein and claims priority therefrom; and is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/651227 filed 29-Aug-03, U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, U.S. Patent Application Serial No. 10/345201 filed 16-Jan-03, U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable

Group of Novel Regulatory Genes and Uses Thereof"; U.S. Patent Application Serial No. 10/651227, filed 29-Aug-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation of U.S. Patent Application Serial No. 10/310914, filed 06-Dec-02, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof", the disclosure of which is hereby incorporated herein and claims priority therefrom; and is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/649653 filed 28-Aug-03, U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03, U.S. Patent Application Serial No. 10/345201 filed 16-Jan-03, U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses

Thereof"; U.S Patent Application Serial No. 10/649653, filed 28-Aug-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation of U.S Patent Application Serial No. 10/321503, filed 18-Dec-02, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof ", the disclosure of which is hereby incorporated herein and claims priority therefrom; and is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/604926 filed 27-Aug-03, U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03, U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S Patent Application Serial No. 10/604926, filed 27-Aug-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses

Thereof" is a continuation of U.S Patent Application Serial No. 10/345201, filed 16-Jan-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" the disclosure of which is hereby incorporated herein and claims priority therefrom; and is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/604726 filed 13-Aug-03, U.S. Patent Application Serial No. 10/604727 filed 13-Aug-03, U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03, U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S Patent Applications Serial Nos.10/604726, filed 13-Aug-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation of U.S Patent Application Serial No.10/293338, filed 14-Nov-02, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and

Uses Thereof ", the disclosure of which is hereby incorporated herein and claims priority therefrom; and is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Provisional Patent Application Serial No. 60/468251 filed 07-May-03, U.S. Patent Application Serial No. 10/345201 filed 16-Jan-03, U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, and U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S Patent Applications Serial Nos.10/604727, filed 13-Aug-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation of U.S Patent Application Serial No.10/293338, filed 14-Nov-02, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof ", the disclosure of which is hereby incorporated herein and claims priority therefrom; and is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Pro-

visional Patent Application Serial No. 60/468251 filed 07-May-03, U.S. Patent Application Serial No. 10/345201 filed 16-Jan-03, U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, and U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02. All of the afore-said patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S. Provisional Patent Application Serial No. 60/468251, filed 07-May-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/345201 filed 16-Jan-03, U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S. Patent Application Serial No. 10/345201, filed 16-Jan-03, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses

Thereof" is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/321503 filed 18-Dec-02, U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. All of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S. Patent Application Serial No. 10/321503, filed 18-Dec-02, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continuation in part of and claims priority from the following patent applications, the disclosures of which applications are all hereby incorporated herein by reference: U.S. Patent Application Serial No. 10/310914 filed 06-Dec-02, and U.S. Patent Application Serial No. 10/293338 filed 14-Nov-02. Both of the aforesaid patent applications are entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof"; U.S. Patent Application Serial No. 10/310914, filed 06-Dec-02, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof" is a continu-

ation in part of U.S Patent Application Serial No.10/293338, filed 14-Nov-02, entitled "Bioinformatically Detectable Group of Novel Regulatory Genes and Uses Thereof ", the disclosure of which is hereby incorporated by reference and claims priority therefrom.

#### REFERENCES CITED

- [0002] Altschul,S.F., Gish,W., Miller,W., Myers,E.W., and Lipman,D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403-410.
- [0003] Dan Gusfield, Algorithms on strings, trees, and sequences: computer science and computational biology, Cambridge University Press, 1997.
- [0004] Elbashir,S.M., Lendeckel,W., and Tuschl,T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. 15, 188-200.
- [0005] Gussow,D. and Clackson,T. (1989). Direct clone characterization from plaques and colonies by the polymerase chain reaction. Nucleic Acids Res. 17, 4000.
- [0006] Hamosh A, Scott AF, Amberger J, Bocchini C, Valle D and McKusick VA.(2002).Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders.Nucleic Acids Res. 30: 52-55.
- [0007] Jenuth,J.P. (2000). The NCBI. Publicly available tools and



resources on the Web. *Methods Mol. Biol.* 132, 301–312.

- [0008] Kirkness, E.F. and Kerlavage, A.R. (1997). The TIGR human cDNA database. *Methods Mol. Biol.* 69, 261–268.
- [0009] Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science* 294, 853–858.
- [0010] Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294, 858–862.
- [0011] Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288, 911–940.
- [0012] Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403, 901–906.
- [0013] Southern, E.M. (1992). Detection of specific sequences among DNA fragments separated by gel electrophoresis. 1975. *Biotechnology* 24, 122–139.
- [0014] Tom M. Mitchell, *Machine Learning*, McGraw Hill, 1997.
- [0015] Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscrip-

tional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. Cell 75, 855–862.

## **BACKGROUND OF THE INVENTION**

### **FIELD OF THE INVENTION**

[0016] The present invention relates to a group of bioinformatically detectable novel human oligonucleotides, here identified as "Genomic Address Messenger" (GAM) oligonucleotides.

[0017] All of abovementioned oligonucleotides are believed to be related to the microRNA (miRNA) group of oligonucleotides.

### **DESCRIPTION OF PRIOR ART**

[0018] miRNA oligonucleotides are short ~22 nucleotide (nt) long, non-coding, regulatory RNA oligonucleotides that are found in a wide range of species. miRNA oligonucleotides are believed to function as specific gene translation repressors and are sometimes involved in cell differentiation.

[0019] The ability to detect novel miRNA oligonucleotides is limited by the methodologies used to detect such oligonucleotides. All miRNA oligonucleotides identified so far ei-

ther present a visibly discernable whole body phenotype, as do Lin-4 and Let-7 (Wightman,B., Ha,I., and Ruvkun,G., Cell 75: 855-862 (1993); Reinhart et al. Nature 403: 901-906 (2000)), or produce sufficient quantities of RNA so as to be detected by standard molecular biological techniques.

[0020] Ninety-three miRNA oligonucleotides have been discovered in several species (Lau et al., Science 294: 858-862 (2001), Lagos-Quintana et al., Science 294: 853-858 (2001)) by sequencing a limited number of clones (300 by Lau and 100 by Lagos-Quintana) of size-fractionated small segments of RNA. miRNAs that were detected in these studies therefore represent the more prevalent among the miRNA oligonucleotide family and cannot be much rarer than 1% of all small ~20 nt-long RNA oligonucleotides.

[0021] The aforementioned studies provide no basis for the detection of miRNA oligonucleotides which either do not present a visually discernable whole body phenotype, or are rare (e.g. rarer than 0.1% of all of the size-fractionated, ~20 nt-long RNA segments that were expressed in the tissues examined), and therefore do not produce large enough quantities of RNA to be detected by

standard biological techniques.

[0022] The following U.S. Patents relate to bioinformatic detection of genes: U.S Patent No. 348935, entitled "Statistical algorithms for folding and target accessibility prediction and design of nucleic acids", U.S Patent No. 6,369,195, entitled "Prostate-specific gene for diagnosis, prognosis and management of prostate cancer", and U.S Patent No.6,291,666 entitled "Spike tissue-specific promoter", each of which is hereby incorporated by reference herein.

#### **BRIEF DESCRIPTION OF SEQUENCE LISTING, TABLES AND COMPUTER PROGRAM LISTING**

[0023] A sequence listing is attached to the present invention, comprising 10068177 genomic sequences, is contained in a file named SEQ\_LIST.txt (1539268KB, 13-May-04), and is hereby incorporated by reference herein.

[0024] Tables relating to genomic sequences are attached to the present application, appear in 21 files (size, creation date), incorporated herein: TABLE\_1.txt (572 MB, 13-May-04), TABLE\_2\_A.txt (619 MB, 13-May-04), TABLE\_2\_B.txt (619 MB, 13-May-04), TABLE\_2\_C.txt (111 MB, 13-May-04), TABLE\_3.txt (22.1 MB, 13-May-04); TABLE\_4.txt (62.3 MB, 13-May-04), TABLE\_5.txt (27.4 MB, 13-May-04), TABLE\_6\_A.txt (619 MB, 13-May-04), TA-

BLE\_6\_B.txt (50.3 MB, 13-May-04), TABLE\_7\_A.txt (619 MB, 13-May-04), TABLE\_7\_B.txt (571 MB, 13-May-04), TABLE\_8\_A.txt (619 MB, 13-May-04), TABLE\_8\_B.txt (619 MB, 13-May-04), TABLE\_9.txt (10.2MB, 13-May-04), TABLE\_10.txt (123MB, 13-May-04), TABLE\_11.txt (79.8MB, 13-May-04), TABLE\_12.txt (75KB, 13-May-04) , TABLE\_13.txt (285 KB,14-May-04) and TABLE\_14.txt (68KB, 13-May-04) all of which are incorporated by reference herein.

[0025] A computer program listing constructed and operative in accordance with a preferred embodiment of the present invention is enclosed on an electronic medium in computer readable form, and is hereby incorporated by reference herein. The computer program listing is contained in 6 files, the name, sizes and creation date of which are as follows: AUXILARY\_FILES.txt (117K, 14-Nov-03); EDIT\_DISTANCE.txt (144K, 24-Nov-03); FIRST-K.txt (96K, 24-Nov-03); HAIRPIN\_PREDICTION.txt (19K, 25-Mar-04); TWO\_PHASED\_SIDE\_SELECTOR.txt (4K, 14-Nov-03); TWO\_PHASED\_PREDICTOR.txt (74K, 14-Nov-03), and BS\_CODE.txt (118K,11-May-04).

#### **SUMMARY OF THE INVENTION**

[0026] The present invention discloses 122,764 novel human

regulatory microRNA-like (miRNA) oligonucleotides referred to here as Genomic Address Messenger (GAM) oligonucleotides, which GAM oligonucleotides are detectable using a novel bioinformatic approach, and go undetected by conventional molecular biology methods. Each GAM oligonucleotide specifically inhibits translation of one of more target genes by hybridization of an RNA transcript encoded by the GAM, to a site located in an untranslated region (UTR) of the mRNA of one or more of the target genes. Also disclosed are 18,602 novel microRNA-cluster like polynucleotides, referred to here as Genomic Record (GR) polynucleotides.

[0027] Accordingly, the invention provides several substantially pure nucleic acids (e.g., genomic DNA, cDNA or synthetic DNA) each comprising a novel human GAM oligonucleotide, vectors comprising the DNAs, probes comprising the DNAs, a method and system for bioinformatic detection of GAM oligonucleotides and their respective targets, laboratory methods for validating expression of GAM oligonucleotides, and a method and system for selectively modulating translation of known target genes of the GAM oligonucleotides.

[0028] The present invention represents a scientific break-

through, disclosing novel miRNA-like oligonucleotides the number of which is dramatically larger than previously believed existed. Prior-art studies reporting miRNA oligonucleotides ((Lau et al., Science 294:858–862 (2001), Lagos-Quintana et al., Science 294: 853–858 (2001)) discovered 93 miRNA oligonucleotides in several species, including 21 in human, using conventional molecular biology methods, such as cloning and sequencing.

[0029] Molecular biology methodologies employed by these studies are limited in their ability to detect rare miRNA oligonucleotides, since these studies relied on sequencing of a limited number of clones (300 clones by Lau and 100 clones by Lagos-Quintana) of small segments (i.e. size-fractionated) of RNA. miRNA oligonucleotides detected in these studies therefore, represent the more prevalent among the miRNA oligonucleotide family, and are typically not be much rarer than 1% of all small ~20 nt-long RNA oligonucleotides present in the tissue from the RNA was extracted.

[0030] Recent studies state the number of miRNA oligonucleotides to be limited, and describe the limited sensitivity of available methods for detection of miRNA oligonucleotides: "The estimate of 255 human miRNA oligonu-

cleotides is an upper bound implying that no more than 40 miRNA oligonucleotides remain to be identified in mammals" (Lim et al., Science, 299:1540 (2003)); "Estimates place the total number of vertebrate miRNA genes at about 200–250" (Ambros et al. Curr. Biol. 13:807–818 (2003)); and "Confirmation of very low abundance miRNAs awaits the application of detection methods more sensitive than Northern blots" (Ambros et al. Curr. Biol. 13:807–818 (2003)).

[0031] The oligonucleotides of the present invention represent a revolutionary new dimension of genomics and of biology: a dimension comprising a huge number of non-protein-coding oligonucleotides which modulate expression of thousands of proteins and are associated with numerous major diseases. This new dimension disclosed by the present invention dismantles a central dogma that has dominated life-sciences during the past 50 years, a dogma which has emphasized the importance of protein-coding regions of the genome, holding non-protein-coding regions to be of little consequence, often dubbing them "junk DNA".

[0032] Indeed, only in November, 2003 has this long held belief as to the low importance of non-protein-coding regions



been vocally challenged. As an example, an article titled "The Unseen Genome – Gems in the Junk" (Gibbs, W.W. Sci. Am. 289:46–53 (2003)) asserts that the failure to recognize the importance of non–protein– coding regions "may well go down as one of the biggest mistakes in the history of molecular biology". Gibbs further asserts that "what was damned as junk because it was not understood, may in fact turn out to be the very basis of human complexity. The present invention provides a dramatic leap in understanding specific important roles of non–protein–coding regions.

[0033] An additional scientific breakthrough of the present invention is a novel conceptual model disclosed by the present invention, which conceptual model is preferably used to encode in a genome the determination of cell–differentiation, utilizing oligonucleotides and polynucleotides of the present invention.

[0034] Using the bioinformatic engine of the present invention, 122,764 GAM oligonucleotides and their respective precursors and targets have been detected. These bioinformatic predictions are supported by robust biological studies. Microarray experiments validated expression of 2,534 GAM oligonucleotides out of a sample of 8,244 tested. Of

these, 1,114 GAM oligonucleotides scored extremely highly: over six standard deviations higher than the background noise of the microarray, and over two standard deviations above their individual mismatch control probes. Thirty eight GAM oligonucleotides were sequenced.

[0035] In various preferred embodiments, the present invention seeks to provide an improved method and system for specific modulation of the expression of specific target genes involved in significant human diseases. It also provides an improved method and system for detection of the expression of novel oligonucleotides of the present invention, which modulate these target genes. In many cases, the target genes may be known and fully characterized, however in alternative embodiments of the present invention, unknown or less well characterized genes may be targeted.

[0036] A "Nucleic acid" is defined as a ribonucleic acid (RNA) molecule, or a deoxyribonucleic acid (DNA) molecule, or complementary deoxyribonucleic acid (cDNA), comprising either naturally occurring nucleotides or non-naturally occurring nucleotides.

[0037] "Substantially pure nucleic acid", "Isolated Nucleic Acid", "Isolated Oligonucleotide" and "Isolated Polynucleotide" are

defined as a nucleic acid that is free of the genome of the organism from which the nucleic acid is derived, and include, for example, a recombinant nucleic acid which is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic nucleic acid of a prokaryote or eukaryote at a site other than its natural site; or which exists as a separate molecule (e.g., a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other nucleic acids.

[0038] An "Oligonucleotide" is defined as a nucleic acid comprising 2–139 nts, or preferably 16–120 nts. A "Polynucleotide" is defined as a nucleic acid comprising 140–5000 nts, or preferably 140–1000 nts.

[0039] A "Complementary" sequence is defined as a first nucleotide sequence which reverses complementary of a second nucleotide sequence: the first nucleotide sequence is reversed relative to a second nucleotide sequence, and wherein each nucleotide in the first nucleotide sequence is complementary to a corresponding nucleotide in the second nucleotide sequence (e.g. ATGGC is the complementary sequence of GCCAT).

[0040] "Hybridization", "Binding" and "Annealing" are defined as

hybridization, under in-vivo physiological conditions, of a first nucleic acid to a second nucleic acid, which second nucleic acid is at least partially complementary to the first nucleic acid.

[0041] A "Hairpin Structure" is defined as an oligonucleotide having a nucleotide sequence that is 50–140 nts in length, the first half of which nucleotide sequence is at least partially complementary to the second part thereof, thereby causing the nucleic acid to fold onto itself, forming a secondary hairpin structure.

[0042] A "Hairpin-Shaped Precursor" is defined as a Hairpin Structure which is processed by a Dicer enzyme complex, yielding an oligonucleotide which is about 19 to about 24 nts in length.

[0043] "Inhibiting translation" is defined as the ability to prevent synthesis of a specific protein encoded by a respective gene by means of inhibiting the translation of the mRNA of this gene. For example, inhibiting translation may include the following steps: (1) a DNA segment encodes an RNA, the first half of whose sequence is partially complementary to the second half thereof; (2) the precursor folds onto itself forming a hairpin-shaped precursor; (3) a Dicer enzyme complex cuts the hairpin-shaped precursor yield-

ing an oligonucleotide that is approximately 22 nt in length; (4) the oligonucleotide binds complementarily to at least one binding site, having a nucleotide sequence that is at least partially complementary to the oligonucleotide, which binding site is located in the mRNA of a target gene, preferably in the untranslated region (UTR) of a target gene, such that the binding inhibits translation of the target protein.

[0044] A "Translation inhibitor site" is defined as the minimal nucleotide sequence sufficient to inhibit translation.

[0045] The present invention describes novel miRNA oligonucleotides, detected using a bioinformatic engine described hereinabove. The ability of this detection engine has been demonstrated using stringent algorithmic criteria, showing that the engine has both high sensitivity, indicated by the high detection rate of published miRNAs and their targets, as well as high specificity, indicated by the low amount of "background" hairpin candidates passing its filters. Laboratory tests, based both on sequencing of predicted miRNA oligonucleotides and on microarray experiments, validated 2534 of the miRNA oligonucleotides in the present invention. Further, at least one of these validated miRNA oligonucleotides binds to 1953 of the 2031

target genes described in the present invention.

[0046] There is thus provided in accordance with a preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1-380 and 6894883 - 7033873.

[0047] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene selected from the group consisting of genes shown in Table 12, Row 1, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1-380 and

6894883 –7033873.

[0048] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide having a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1–380 and 6894883 –7033873.

[0049] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a portion of a mRNA transcript of a target gene, and anneals to a second oligonucleotide that is endogenously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses expression of the target gene, and wherein nucleotide sequence of the second nucleotide is selected from the group consisting of SEQ ID NOs: 1–380 and 6894883 –7033873.

[0050] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable first oligonucleotide which is a portion of a mRNA transcript of a target gene selected from the group consisting of genes shown in Table 12 row 1, and anneals to a second oligonucleotide that is endoge-

nously processed from a hairpin precursor, wherein binding of the first oligonucleotide to the second oligonucleotide represses expression of the target gene, and wherein nucleotide sequence of the second nucleotide is selected from the group consisting of SEQ ID NOs: 1-380 and 6894883 -7033873.

[0051] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable oligonucleotide having a nucleotide sequence selected from the group consisting of SEQ ID NOs: 5054808-6757247.

[0052] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Multiple Sclerosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 2.

[0053] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinform-



matically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Alzheimer, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 3.

[0054] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Prostate cancer, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 4.

[0055] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Respiratory Syncytial Virus, wherein bind-

ing of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 5.

[0056] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Inflammatory Bowel Diseases, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 6.

[0057] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Chronic obstructive pulmonary disease, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence

identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 7.

[0058] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Myasthenia Gravis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 8.

[0059] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Nephrogenic diabetes insipidus, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 9.

[0060] There is still further provided in accordance with another

preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Carcinoid, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 10.

[0061] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Esophageal cancer, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 11.

[0062] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene

associated with Polyposis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 12.

[0063] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Allergic contact dermatitis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 13.

[0064] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Myopathy, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide

has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 14.

[0065] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Otitis Media, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 15.

[0066] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Lung cancer, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 16.

[0067] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Enterovirus, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 18.

[0068] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Stroke, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 19.

[0069] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which an-

neals to a portion of a mRNA transcript of a target gene associated with Hodgkin Disease, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 20.

[0070] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Amyloidosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 21.

[0071] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Depressive Disorder, wherein binding of the oligonucleotide to the mRNA transcript represses ex-



pression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 22.

[0072] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Clostridium, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 23.

[0073] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with HIV, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown

in Table 14 row 24.

[0074] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Ventricular Fibrillation, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 25.

[0075] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Hyperlipidemia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 26.

[0076] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Atherosclerosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 27.

matically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Lymphoma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 27.

[0077] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Atopic dermatitis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 28.

[0078] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Pagets Disease, wherein binding of the

oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 29.

[0079] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Emphysema, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 30.

[0080] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Ventricular tachycardia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nu-

cleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 31.

[0081] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Hepatocellular carcinoma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 32.

[0082] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Kidney Failure, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 33.

[0083] There is further provided in accordance with another pre-

ferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Addisons disease, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 34.

[0084] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Herpes, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 35.

[0085] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene

associated with Malaria, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 36.

[0086] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Breast cancer, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 37.

[0087] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Leukemia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide

has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 38.

[0088] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Alopecia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 39.

[0089] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Hepatitis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 40.



[0090] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Cataract, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 41.

[0091] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Encephalitis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 42.

[0092] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which an-

neals to a portion of a mRNA transcript of a target gene associated with Cholestasis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 43.

[0093] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Schizophrenia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 44.

[0094] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Hyperglycemia, wherein binding of the oligonucleotide to the mRNA transcript represses expres-

sion of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 45.

[0095] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Megaloblastic anemia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 46.

[0096] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Endometrial carcinoma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of

SEQ ID NOs shown in Table 14 row 47.

[0097] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Burkitt lymphoma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 48.

[0098] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Crohn disease, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 49.

[0099] There is further provided in accordance with another preferred embodiment of the present invention a bioinform-

matically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Osteoarthritis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 50.

[0100] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Pancreatitis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 51.

[0101] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Fragile X Syndrome, wherein binding of

the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 52.

[0102] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Anorexia Nervosa, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 53.

[0103] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Bladder cancer, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide se-

quence selected from the group consisting of SEQ ID NOs shown in Table 14 row 54.

[0104] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Insulin-Dependent Diabetes Mellitus, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 55.

[0105] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Sideroblastic anemia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 56.

[0106] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Celiac Disease, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 57.

[0107] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Diabetes Mellitus, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 58.

[0108] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which an-



neals to a portion of a mRNA transcript of a target gene associated with Basal cell carcinoma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 59.

[0109] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Cytomegalovirus, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 60.

[0110] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Aids, wherein binding of the oligonucleotide to the mRNA transcript represses expression of

the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 61.

[0111] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Small cell carcinoma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 62.

[0112] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Diabetic Nephropathy, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of

SEQ ID NOs shown in Table 14 row 63.

[0113] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Adrenal cortical carcinoma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 65.

[0114] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Toxoplasmosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 66.

[0115] There is further provided in accordance with another preferred embodiment of the present invention a bioinform-

matically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Bundle-Branch Block, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 67.

[0116] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Thyroiditis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 68.

[0117] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Urethral neoplasms, wherein binding of

the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 69.

[0118] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Adenovirus, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 70.

[0119] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Atherosclerosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide se-

quence selected from the group consisting of SEQ ID NOs shown in Table 14 row 71.

[0120] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Infectious Mononucleosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 72.

[0121] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Non-Insulin-Dependent Diabetes Mellitus, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 73.

[0122] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Virus Diseases, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 74.

[0123] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Hypertrophic cardiomyopathy, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 75.

[0124] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which an-

neals to a portion of a mRNA transcript of a target gene associated with Syphilis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 76.

[0125] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Thrombocytopenia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 77.

[0126] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Cerebrovascular Accident, wherein binding of the oligonucleotide to the mRNA transcript re-



presses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 78.

[0127] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Skin Neoplasms, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 79.

[0128] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Cleft Palate, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs

shown in Table 14 row 80.

[0129] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Obesity, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 81.

[0130] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Picornaviridae, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 82.

[0131] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Influenza, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 83.

matically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Nonsmall cell lung cancer, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 83.

[0132] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Dermatomyositis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 84.

[0133] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Migraine, wherein binding of the oligonu-

cleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 85.

[0134] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Meningitis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 86.

[0135] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Renal Tubular Acidosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nu-

cleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 87.

[0136] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Pancreatic cancer, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 88.

[0137] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Ulcerative colitis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 89.

[0138] There is moreover provided in accordance with another

preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Epilepsy, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 90.

[0139] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Cholelithiasis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 91.

[0140] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene

associated with Intestinal Neoplasms, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 92.

[0141] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Renal cell carcinoma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 93.

[0142] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Cirrhosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at

least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 94.

[0143] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Peritonitis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 95.

[0144] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Appendicitis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 96.



[0145] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Papilloma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 97.

[0146] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Down Syndrome, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 98.

[0147] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which an-

neals to a portion of a mRNA transcript of a target gene associated with Nephrolithiasis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 99.

[0148] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Aortic Aneurysm, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 100.

[0149] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Vascular dementia, wherein binding of the oligonucleotide to the mRNA transcript represses expres-

sion of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 101.

[0150] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Infertility, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 102.

[0151] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Thyroid carcinoma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs

shown in Table 14 row 103.

[0152] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Thrombosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 104.

[0153] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Asthma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 105.

[0154] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Diabetes Mellitus, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 106.

matically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Diverticulitis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 106.

[0155] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Tuberculosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 108.

[0156] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Multiinfarct dementia, wherein binding of

the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 109.

[0157] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Cervical cancer, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 110.

[0158] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Beta Thalassemia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide se-

quence selected from the group consisting of SEQ ID NOs shown in Table 14 row 111.

[0159] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Hepatocellular carcinoma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 112.

[0160] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Psoriasis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 113.

[0161] There is additionally provided in accordance with another

preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Diphtheria, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 114.

[0162] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Bronchiectasis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 115.

[0163] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene



associated with EBV, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 116.

[0164] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Coronary disease, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 117.

[0165] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Polyposis coli, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide

has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 118.

[0166] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Influenza, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 119.

[0167] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Parkinson, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 120.

[0168] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Hemolytic anemia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 121.

[0169] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Medullary thyroid carcinoma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 122.

[0170] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which an-

neals to a portion of a mRNA transcript of a target gene associated with Sickle cell anemia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 123.

[0171] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Deafness, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 124.

[0172] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Diabetic Neuropathies, wherein binding of the oligonucleotide to the mRNA transcript represses ex-

pression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 125.

[0173] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Psoriatic arthritis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 126.

[0174] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Barrett Esophagus, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs

shown in Table 14 row 127.

[0175] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Cerebral Hemorrhage, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 128.

[0176] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Cerebral Infarction, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 129.

[0177] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Cerebral Hemorrhage, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 130.

matically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with E.coli, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 130.

[0178] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Urticaria, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 131.

[0179] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Attention Deficit Disorder, wherein bind-

ing of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 132.

[0180] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Pituitary tumor, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 133.

[0181] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Enuresis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence



selected from the group consisting of SEQ ID NOs shown in Table 14 row 134.

[0182] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Osteoporosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 135.

[0183] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Urinary calculi, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 136.

[0184] There is still further provided in accordance with another

preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Multiple Myeloma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 137.

[0185] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Aplastic anemia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 138.

[0186] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene

associated with Gestational Diabetes, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 139.

[0187] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Rheumatoid arthritis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 140.

[0188] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Duodenal Neoplasms, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonu-

cleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 141.

[0189] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Hypertrophic Cardiomyopathy, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 142.

[0190] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Myocardial Infarction, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 143.

[0191] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Left Ventricular Dysfunction, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 144.

[0192] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Postpartum depression, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 145.

[0193] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which an-

neals to a portion of a mRNA transcript of a target gene associated with Colorectal cancer, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 146.

[0194] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Transitional cell carcinoma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 147.

[0195] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Alpha thalassemia, wherein binding of the oligonucleotide to the mRNA transcript represses expres-

sion of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 148.

[0196] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Cleft Lip, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 149.

[0197] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Hypercholesterolemia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of

SEQ ID NOs shown in Table 14 row 150.

[0198] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Sudden cardiac death, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 151.

[0199] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Atrial fibrillation, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 152.

[0200] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Atrial fibrillation, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 153.



matically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Hypertension, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 153.

[0201] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Ovarian cancer, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 154.

[0202] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Coronary spasm, wherein binding of the

oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 155.

[0203] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Hemophilia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 157.

[0204] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Peripheral Vascular Diseases, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a

nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 158.

[0205] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Bacillary Dysentery, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 159.

[0206] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Macular Degeneration, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 160.

[0207] There is further provided in accordance with another pre-

ferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Mycobacterium, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 161.

[0208] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Cushing Syndrome, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 162.

[0209] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene

associated with Melanoma, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 163.

[0210] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Bipolar Disorder, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 164.

[0211] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Coronary artery disease, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonu-

cleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 166.

[0212] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Dementia, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 167.

[0213] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Lupus Erythematosus, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 168.

[0214] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Rhinitis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 169.

[0215] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Peptic Ulcer, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 170.

[0216] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which an-

neals to a portion of a mRNA transcript of a target gene associated with Cystic fibrosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 171.

[0217] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Autism, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 172.

[0218] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with HTLV, wherein binding of the oligonucleotide to the mRNA transcript represses expression of



the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 173.

[0219] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Sinusitis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 174.

[0220] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Diabetic Retinopathy, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of

SEQ ID NOs shown in Table 14 row 176.

[0221] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Antisocial Personality Disorder, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 177.

[0222] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which anneals to a portion of a mRNA transcript of a target gene associated with Amyotrophic Lateral Sclerosis, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the oligonucleotide has at least 80% sequence identity with a nucleotide sequence selected from the group consisting of SEQ ID NOs shown in Table 14 row 178.

[0223] There is further provided in accordance with another preferred embodiment of the present invention a method for

treatment of a disease involving a tissue in which a protein is pathologically expressed to an undesirable extent, the protein having a messenger RNA, the method including: providing a material which modulates activity of a microRNA oligonucleotide which binds complementarily to a segment of the messenger RNA, and introducing the material into the tissue, causing modulation of the activity of the microRNA oligonucleotide and thereby modulating expression of the protein in a desired manner.

[0224] There is still further provided in accordance with another preferred embodiment of the present invention a method for treatment of a disease involving tissue in which a protein is pathologically expressed to an undesirable extent, the protein having a messenger RNA, the method including: providing a material which at least partially binds a segment of the messenger RNA that is bound complementarily by a microRNA oligonucleotide, thereby modulating expression of the protein, and introducing the material into the tissue, thereby modulating expression of the protein.

[0225] There is additionally provided in accordance with another preferred embodiment of the present invention a method for treatment of a disease involving a tissue in which a

protein is pathologically over-expressed, the protein having a messenger RNA, the method including: providing a microRNA oligonucleotide which binds complementarily to a segment of the messenger RNA, and introducing the microRNA oligonucleotide into the tissue, causing the microRNA oligonucleotide to bind complementarily to a segment of the messenger RNA and thereby inhibit expression of the protein.

[0226] There is moreover provided in accordance with another preferred embodiment of the present invention a method for treatment of a disease involving a tissue in which a protein is pathologically over-expressed, the protein having a messenger RNA, the method including: providing a chemically-modified microRNA oligonucleotide which binds complementarily to a segment of the messenger RNA, and introducing the chemically-modified microRNA oligonucleotide into the tissue, causing the microRNA oligonucleotide to bind complementarily to a segment of the messenger RNA and thereby inhibit expression of the protein.

[0227] There is further provided in accordance with another preferred embodiment of the present invention a method for treatment of a disease involving a tissue in which a pro-

tein is pathologically under-expressed, the protein having a messenger RNA, the method including: providing an oligonucleotide that inhibits activity of a microRNA oligonucleotide which binds complementarily to a segment of the messenger RNA, and introducing the oligonucleotide into the tissue, causing inhibition of the activity of the microRNA oligonucleotide and thereby promotion of translation of the protein.

[0228] There is still further provided in accordance with another preferred embodiment of the present invention a method for treatment of a disease involving a tissue in which a protein is pathologically under-expressed, the protein having a messenger RNA, the method including: providing a chemically-modified oligonucleotide that inhibits activity of a microRNA oligonucleotide which binds complementarily to a segment of the messenger RNA, and introducing the chemically-modified oligonucleotide into the tissue, causing inhibition of the activity of the microRNA oligonucleotide and thereby promotion of translation of the protein.

[0229] There is additionally provided in accordance with another preferred embodiment of the present invention a method for diagnosis of a disease involving a tissue in which a

protein is expressed to abnormal extent, the protein having a messenger RNA, the method including: assaying a microRNA oligonucleotide which at least partially binds a segment of the messenger RNA and modulates the expression of the protein, thereby providing an indication of at least one parameter of the disease.

[0230] There is moreover provided in accordance with another preferred embodiment of the present invention a method for detection of expression of an oligonucleotide, the method including: determining a first nucleotide sequence of a first oligonucleotide, which first nucleotide sequence is not complementary to a genome of an organism, receiving a second nucleotide sequence of a second oligonucleotide whose expression is sought to be detected, designing a third nucleotide sequence that is complementary to the second nucleotide sequence of the second oligonucleotide, and a fourth nucleotide sequence that is complementary to a fifth nucleotide sequence which is different from the second nucleotide sequence of the second oligonucleotide by at least one nucleotide, synthesizing a first oligonucleotide probe having a sixth nucleotide sequence including the third nucleotide sequence followed by the first nucleotide sequence of the

first oligonucleotide, and a second oligonucleotide probe having a seventh nucleotide sequence including the fourth nucleotide sequence followed by the first nucleotide sequence of the first oligonucleotide, locating the first oligonucleotide probe and the second oligonucleotide probe on a microarray platform, receiving an RNA test sample from at least one tissue of the organism, obtaining size-fractionated RNA from the RNA test sample, amplifying the size-fractionated RNA, hybridizing the adaptor-linked RNA with the first and second oligonucleotide probes on the microarray platform, and determining expression of the first oligonucleotide in the at least one tissue of the organism, based at least in part on the hybridizing.

[0231] There is further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated polynucleotide which is endogenously processed into a plurality of hairpin-shaped precursor oligonucleotides, each of which is endogenously processed into a respective oligonucleotide, which in turn anneals to a portion of a mRNA transcript of a target gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene.

[0232] There is still further provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein the target gene does not encode a protein.

[0233] There is additionally provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene, wherein binding of the oligonucleotide to the mRNA transcript represses expression of the target gene, and wherein a function of the oligonucleotide includes modulation of cell type.

[0234] There is moreover provided in accordance with another preferred embodiment of the present invention a bioinformatically detectable isolated oligonucleotide which is endogenously processed from a hairpin-shaped precursor, and anneals to a portion of a mRNA transcript of a target gene, wherein binding of the oligonucleotide to the mRNA



transcript represses expression of the target gene, and wherein the oligonucleotide is maternally transferred by a cell to at least one daughter cell of the cell, and a function of the oligonucleotide includes modulation of cell type of the daughter cell.

[0235] There is further provided in accordance with another preferred embodiment of the present invention a method for bioinformatic detection of microRNA oligonucleotides, the method including: bioinformatically detecting a hairpin-shaped precursor oligonucleotide, bioinformatically detecting an oligonucleotide which is endogenously processed from the hairpin-shaped precursor oligonucleotide, and bioinformatically detecting a target gene of the oligonucleotide wherein the oligonucleotide anneals to at least one portion of a mRNA transcript of the target gene, and wherein the binding represses expression of the target gene, and the target gene is associated with a disease.

#### **BRIEF DESCRIPTION OF DRAWINGS**

[0236] Fig. 1 is a simplified diagram illustrating a genomic differentiation enigma that the present invention addresses;

[0237] Figs. 2, 3 and 4 are schematic diagrams which, when taken together, provide an analogy that illustrates a con-

ceptual model of the present invention, addressing the genomic differentiation enigma;

[0238] Figs. 5A and 5B are schematic diagrams which, when taken together, illustrate a "genomic records" concept of the conceptual model of the present invention, addressing the genomic differentiation enigma;

[0239] Fig. 6 is a schematic diagram illustrating a "genomically programmed cell differentiation" concept of the conceptual model of the present invention, addressing the genomic differentiation enigma;

[0240] Fig. 7 is a schematic diagram illustrating a "genomically programmed cell-specific protein expression modulation" concept of the conceptual model of the present invention, addressing the genomic differentiation enigma;

[0241] Fig. 8 is a simplified diagram illustrating a mode by which an oligonucleotide of a novel group of oligonucleotides of the present invention modulates expression of known target genes;

[0242] Fig. 9 is a simplified block diagram illustrating a bioinformatic oligonucleotide detection system capable of detecting oligonucleotides of the novel group of oligonucleotides of the present invention, which system is constructed and operative in accordance with a preferred em-

bodiment of the present invention;

[0243] Fig. 10 is a simplified flowchart illustrating operation of a mechanism for training of a computer system to recognize the novel oligonucleotides of the present invention, which mechanism is constructed and operative in accordance with a preferred embodiment of the present invention;

[0244] Fig. 11A is a simplified block diagram of a non-coding genomic sequence detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0245] Fig. 11B is a simplified flowchart illustrating operation of a non-coding genomic sequence detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0246] Fig. 12A is a simplified block diagram of a hairpin detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0247] Fig. 12B is a simplified flowchart illustrating operation of a hairpin detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0248] Fig. 13A is a simplified block diagram of a Dicer-cut location detector constructed and operative in accordance

with a preferred embodiment of the present invention;

[0249] Fig. 13B is a simplified flowchart illustrating training of a Dicer-cut location detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0250] Fig. 13C is a simplified flowchart illustrating operation of a Dicer-cut location detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0251] Fig. 14A is a simplified block diagram of a target gene binding site detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0252] Fig. 14B is a simplified flowchart illustrating operation of a target gene binding site detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0253] Fig. 15 is a simplified flowchart illustrating operation of a function and utility analyzer constructed and operative in accordance with a preferred embodiment of the present invention;

[0254] Fig. 16 is a simplified diagram describing a novel bioinformatically-detected group of regulatory polynucleotides,

referred to here as Genomic Record (GR) polynucleotides, each of which encodes an "operon-like" cluster of novel miRNA-like oligonucleotides, which in turn modulate expression of one or more target genes;

[0255] Fig. 17 is a simplified diagram illustrating a mode by which human oligonucleotides of a novel group of operon-like polynucleotides of the present invention, modulate expression of other such polynucleotides, in a cascading manner;

[0256] Fig. 18 is a block diagram illustrating an overview of a methodology for finding novel human oligonucleotides and novel operon-like human polynucleotides of the present invention, and their respective functions;

[0257] Fig. 19 is a block diagram illustrating different utilities of novel oligonucleotides and novel operon-like polynucleotides, both of the present invention;

[0258] Figs. 20A and 20B are simplified diagrams which, when taken together, illustrate a mode of oligonucleotide therapy applicable to novel oligonucleotides of the present invention;

[0259] Fig. 21A is a bar graph illustrating performance results of a hairpin detector constructed and operative in accordance with a preferred embodiment of the present inven-

tion;

[0260] Fig. 21B is a line graph illustrating accuracy of a Dicer-cut location detector constructed and operative in accordance with a preferred embodiment of the present invention;

[0261] Fig. 21C is a bar graph illustrating performance results of the target gene binding site detector 118, constructed and operative in accordance with a preferred embodiment of the present invention.

[0262] Fig. 22 is a summary table of laboratory results validating expression of novel human oligonucleotides detected by a bioinformatic oligonucleotide detection engine constructed and operative in accordance with a preferred embodiment of the present invention, thereby validating its efficacy;

[0263] Fig. 23A is a schematic representation of an "operon-like" cluster of novel human hairpin sequences detected by a bioinformatic oligonucleotide detection engine constructed and operative in accordance with a preferred embodiment of the present invention, and non-GAM hairpin sequences used as negative controls thereto;

[0264] Fig. 23B is a schematic representation of secondary folding of hairpins of the operon-like cluster of Fig. 23A;

[0265] Fig. 23C is a photograph of laboratory results demon-

strating expression of novel oligonucleotides of Figs. 23A and 23B and lack of expression of the negative controls, thereby validating efficacy of bioinformatic detection of GAM oligonucleotides and GR polynucleotides detected by a bioinformatic oligonucleotide detection engine, constructed and operative in accordance with a preferred embodiment of the present invention;

[0266] Fig. 24A is an annotated sequence of EST72223 comprising known human miRNA oligonucleotide MIR98 and novel human oligonucleotide GAM25 PRECURSOR detected by the oligonucleotide detection system of the present invention; and

[0267] Figs. 24B, 24C and 24D are pictures of laboratory results demonstrating laboratory confirmation of expression of known human oligonucleotide MIR98 and of novel bioinformatically-detected human GAM25 RNA respectively, both of Fig. 24A, thus validating the bioinformatic oligonucleotide detection system of the present invention;

[0268] Fig. 25A, 25B and 25C are schematic diagrams which, when taken together, represent methods of designing primers to identify specific hairpin oligonucleotides in accordance with a preferred embodiment of the present invention.

- [0269] Fig. 26A is a simplified flowchart illustrating construction of a microarray constructed and operative to identify novel oligonucleotides of the present invention, in accordance with a preferred embodiment of the present invention;
- [0270] Fig. 26B is a simplified block diagram illustrating design of a microarray constructed and operative to identify novel oligonucleotides of the present invention, in accordance with a preferred embodiment of the present invention;
- [0271] Fig. 26C is a flowchart illustrating a mode of preparation and amplification of a cDNA library in accordance with a preferred embodiment of the present invention;
- [0272] Fig. 27A is a line graph showing results of detection of known microRNA oligonucleotides and of novel GAM oligonucleotides, using a microarray constructed and operative in accordance with a preferred embodiment of the present invention;
- [0273] Fig. 27B is a line graph showing specificity of hybridization of a microarray constructed and operative in accordance with a preferred embodiment of the present invention; and
- [0274] Fig. 27C is a summary table demonstrating detection of known microRNA oligonucleotides using a microarray constructed and operative in accordance with a preferred



embodiment of the present invention.

#### **BRIEF DESCRIPTION OF SEQUENCES**

[0275] A Sequence Listing of genomic sequences of the present invention designated SEQ ID NO:1 through SEQ ID: 10068177 is attached to this application, and is hereby incorporated herein. The genomic listing comprises the following nucleotide sequences: nucleotide sequences of 122764 GAM precursors of respective novel oligonucleotides of the present invention; nucleotide sequences of 139368 GAM RNA oligonucleotides of respective novel oligonucleotides of the present invention; and nucleotide sequences of 1709460 target gene binding sites of respective novel oligonucleotides of the present invention.

#### **DETAILED DESCRIPTION**

[0276] Reference is now made to Fig. 1, which is a simplified diagram providing a conceptual explanation of a genomic differentiation enigma, which the present invention addresses, inter alia.

[0277] Fig. 1 depicts various types of cells in an organism, such as a cartilage cell designated by reference numeral 1, a liver cell designated by reference numeral 2, a fibroblast cell designated by reference numeral 3, and a bone cell

designated by reference numeral 4, all containing identical DNA designated by reference numeral 5. Notwithstanding that the various types of cells are all derived from a common initial fertilized egg cell designated by reference numeral 6, each of these cells expresses different proteins and accordingly acquire a different shape and function.

[0278] The present invention proposes inter alia that the inevitable conclusion from the foregoing is strikingly simple: the genome must contain a modular differentiation coding system. In other words, the genome of each cell must include multiple modules or records, possibly a different one for each cell type, as well as a mechanism by which each cell at its inception is instructed which one of the multiple records will govern its behavior.

[0279] This modular code concept may be somewhat difficult to grasp, since most persons are accustomed to view things from an external viewpoint. An architect, for example, looks at a plan of a building, which details exactly where each element (block, window, door, electrical switch, etc.) is to be placed relative to all other elements. Using the plan, the architect instructs the builders to place these elements in their designated places. This is an example of

an external viewpoint: the architect is external to the plan, which itself is external with respect to the physical building, and with respect to its various elements. The architect may therefore act as an "external organizing agent" who can see the full picture and the relationships between all of the elements and is able to instruct from the outside where to place each of them.

[0280] According to a preferred embodiment of the present invention, genomic differentiation coding, in contrast to architectural building, functions without any external organizing agent. It comprises a smart block (the first cell), which is the architect and the plan. This smart block continuously duplicates itself, somehow knowing when to manifest itself as a block and when as a window, door, or electrical switch.

[0281] Reference is now made to Figs. 2A-4 which are schematic diagrams which, when taken together, provide an analogy that illustrates a conceptual model of the present invention, which conceptual model addresses the genomic differentiation enigma.

[0282] Reference is now made to Fig. 2A. A hypothetical talented chef, designated by reference numeral 7, is capable of preparing any dish provided that he is given specific writ-

ten cooking instructions. The chef 7 is equipped with two items: (a) a recipe book 8, designated by reference numeral 8, and (b) a small note, designated by reference numeral 9, having a number scribbled on it. The recipe book 8 comprises multiple pages, each page detailing how to prepare a specific dish. The small note 9 indicates the page to be opened, and therefore the dish to be prepared. The chef looks at the page number written on the note, opens the recipe book 8 to the appropriate page, and prepares the dish according to the written instructions on this page. In the example shown in Fig. 2A, the chef 7 is holding a small note 9 bearing the number 127. He therefore opens the recipe book 8 to page 127, designated by reference numeral 10. Since this page contains the recipe for preparing bread, the chef 7 prepares a loaf of bread, designated by reference numeral 12. Pages of the recipe book 8, such as page 127 (designated by reference numeral 10) in the example shown in Fig. 2A, contain additional information, designated by reference numeral 11. The nature of the additional information 11 is further elaborated hereinbelow with reference to Figs. 3 and 4.

[0283] Reference is now made to Fig. 2B, which depicts two identical chefs, a first chef, designated by reference numeral

13, and a second chef, designated by reference numeral 14, both holding an identical recipe book 8. Although the first chef 13 and the second chef 14 are identical and hold identical recipe books 8, they differ in that they hold different small notes. The first chef 13 holds a small note designated by reference numeral 9, having the number 127 written on it, whereas the second chef 14 holds a small note designated by reference numeral 15, having the number 134 written on it. Accordingly, the first chef 13 opens the recipe book 8 to page 127, as designated by reference numeral 10 and, based on the instructions written on page 127 prepares a loaf of bread, designated by reference numeral 12. The second chef 14 opens the recipe book 8 to page 134, as designated by reference numeral 16 and, based on the instructions written on page 134, prepares a pie, designated by reference numeral 17. Pages in the recipe book 8, such as pages 127 and 134 designated by reference numerals 10 and 16 respectively in the examples shown in Fig. 2B, contain additional information, designated by reference numeral 11. The nature of the additional information 11 is further elaborated hereinbelow with reference to Figs. 3 and 4.

[0284] Reference is now made to Fig. 3, which illustrates a mode

by which an imaginary chef can duplicate himself yielding two identical chefs, instructing each of the identical duplicate chefs to prepare a different dish. As an example, Fig. 3 shows a chef, designated by reference numeral 21, duplicating himself to yield two duplicate chefs: a first duplicate chef, designated by reference numeral 22, and a second duplicate chef, designated by reference numeral 23. The duplicate chefs are identical to each other and to chef 21.

[0285] Like chef 7 and chef 13 of Fig. 2A and 2B, Fig. 3 shows chef 21 holding a recipe book 8 and receiving a note 9 bearing the number 127. Chef 21 therefore opens the recipe book 8 to page 127, designated by reference numeral 10, and prepares a loaf of bread 12. However, Fig. 3 also elaborates some of the additional information 11 (Figs. 2A and 2B) found on page 127, designated by reference numeral 10: the bottom of page 127 bears two numbers 134 and 157.

[0286] Chef 21 is trained to perform the following three actions when he is finished preparing a dish: (a) Duplicate himself yielding two duplicate chefs, the first duplicate chef 22 and the second duplicate chef 23; (b) Duplicate his recipe book 8, handing an identical copy to each of the duplicate

chefs 22 and 23; and (c) Write down on each of two notes one of the numbers that is found at the bottom of the page to which he was instructed to open. In the example illustrated by Fig. 3, chef 21 is instructed to open the recipe book 8 to page 127, designated by reference numeral 10, write the numbers 134 and 157 on two respective notes, a first note designated by reference numeral 15 and the second note designated by reference numerals 24. Chef 21 is further trained to hand the first note 15 bearing the number 134, to the first duplicate chef 22 and the second note 24 bearing the number 157, to the second duplicate chef 23.

[0287] Accordingly, the first duplicate chef 22 receives note 15 bearing the number 134 and therefore opens the recipe book 8 to page 134, designated by reference numeral 16, and prepares a pie, designated by reference numeral 17. The second duplicate chef 23 receives note 24 bearing the number 157 and therefore opens the recipe book 8 to page 157, designated by reference numeral 25, and prepares rice, designated by reference numeral 26.

[0288] It is appreciated that while chef 21 and duplicate chefs 22 and 23 are identical and hold identical recipe books 8, they each prepare a different dish. It is also appreciated

that the dishes prepared by the first duplicate chef 22 and the second duplicate chef 23 are determined by chef 21 and are mediated by the differently numbered notes 15 and 24 passed on from chef 21 to duplicate chefs 22 and 23, respectively.

[0289] Further, it is appreciated that the mechanism illustrated by Fig. 3 enables an unlimited lineage of chefs to divide into duplicate, identical chefs and to determine the dishes those duplicate chefs would prepare. As an example, since the first duplicate chef 22 is directed to page 134, as designated by reference numeral 16, when he duplicates himself (not shown), he will instruct his two duplicate chefs to prepare dishes specified on particular pages, the numbers of which are written at the bottom of page 134, i.e. pages 114 and 193, respectively. Similarly, the second duplicate chef 23 will instruct its duplicate chefs to prepare dishes specified on pages 121 and 146, respectively, etc.

[0290] Reference is now made to Fig. 4, which illustrates a mode by which a chef can prepare a dish based on instructions written in a shorthand format: The page to which a chef is directed by a small note he is given merely contains a list of numbers which further direct him to other pages, each



specifying how to prepare an ingredient of the dish to be prepared.

[0291] To illustrate this shorthand format, Fig. 4 shows a chef, designated by reference numeral 27, holding the recipe book 8 and the note 9 which bears the number 127. Chef 27 accordingly opens the recipe book 8 to page 127, designated by reference numeral 10, and based on instructions on this page, prepares bread 12. This is similar to chefs 7, 13 and 21 of Figs. 2A, 2B and 3, respectively.

[0292] However, Fig. 4 also further elaborates on some of the additional information 11 (Figs. 2A and 2B) that is written on page 127, designated by reference numeral 10. The cooking instructions found on page 127, designated by reference numeral 10, for making bread 12 are written in a shorthand format, comprising only three numbers: 118, 175 and 183. Chef 27 writes these numbers on three respective notes designated by reference numerals 28–30. The notes 28–30 are then used to turn to corresponding pages 118, 175 and 183, designated by reference numerals 31–33 of the recipe book 8, which pages provide instructions for the preparation of ingredients required for making bread 12: flour 34, milk 35 and salt 36.

[0293] The analogy provided by Figs. 2A–4 illustrates the con-

ceptual model of the present invention addressing the genomic differentiation enigma, and may be explained as follows: The chefs and duplicate chefs 7, 13, 14, 21-23 and 27 (Figs. 2A-4) in the analogy represent cells. The recipe book 8 represents the DNA 5 (Fig. 1). Preparing dishes such as bread 12, pie 17 or rice 26 (all of Fig. 3) represent the cell manifesting itself as a specific cell type, such as cartilage cell 1, liver cell 2, fibroblast cell 3, or bone cell 4 (all of Fig. 1). Ingredients of a dish, such as flour 34, milk 35 and salt 36 (all of 4), represent proteins typically expressed by a particular cell type, such as 1-4. In the same way that the different chefs of the analogy have the same recipe book 8 yet prepare different dishes, so do different cells in an organism contain the same DNA 5 yet manifest themselves as different cell types 1-4 by expressing proteins typical of these respective cell types. Application of the analogy illustrated in Figs. 2A-4 to the field of cell biology is further described hereinbelow with reference to Figs. 5A-7.

[0294] Reference is now made to Figs. 5A and 5B which are schematic diagrams that, when taken together, illustrate a Genomic Records concept of the present invention, addressing the genomic differentiation enigma. Figs. 5A and

5B correspond to Figs. 2A and 2B of the chef analogy described hereinabove.

[0295] An important aspect of the present invention is the Genomic Records concept. According to a preferred embodiment of the present invention, the DNA (the recipe book 8 in analogy) comprises a very large number of Genomic Records (analogous to pages in the recipe book 8, such as pages 127, 134, and 157, designated by reference numerals 10, 16 and 25, respectively) containing instructions for differentiation of various different cell types or developmental process. Each Genomic Record comprises at least one very short genomic sequence, which functions as a "Genomic Address" of that Genomic Record (analogous to a page number, such as the numbers 127, 134 and 157 (reference numerals 10, 16 and 25) that appear in the recipe book 8 of Fig. 3). At its inception, each cell receives a short RNA segment (analogous to the scribbled short note, such as 9, 15 and 24 of Fig. 3) in addition to the DNA (analogous to the recipe book 8). This short RNA segment binds complementarily to a "Genomic Address" sequence of one of the Genomic Records, thereby modulating expression of that Genomic Record, and, accordingly, determining the cell's fate (analogous to open-

ing the recipe book 8 to a page corresponding to a number on the scribbled note, thereby determining the dish to be prepared). A Genomic Record may also comprise multiple short RNA segments, each of which binds complementarily to a target protein-coding gene, thus modulating expression of this target gene. The Genomic Records concept is analogous to the shorthand format illustrated by Fig. 4 whereby a page, such as page 127, designated by reference numeral 10, points to other pages, such as pages 118, 175 and 183, designated by reference numerals 31-33 respectively, encoding various ingredients, such as flour 34, milk 35 and salt 36, all of Fig. 4.

[0296] Reference is now made to Fig. 5A. Fig. 5A illustrates a cell 37 having a genome 38. The genome 38 comprises a plurality of Genomic Records, some of which correspond to specific cell types. As an example, six such genomic records are shown, corresponding to six cell types: lymphocyte (LYMPH) genomic record 39, fibroblast (FIBRO) genomic record 40, muscle genomic record 41, bone genomic record 42, cartilage (CARTIL.) genomic record 43 and nerve genomic record 44. Each genomic record comprises genomic instructions on differentiation into a specific cell type, as further elaborated hereinbelow with ref-

erence to Fig. 7. At its inception, cell 37 receives a maternal short RNA segment 46, which activates one of the genomic records, causing the cell to differentiate according to the instructions this genomic record comprises. As an example, Fig. 5A illustrates cell 37 reception of a maternal short RNA segment, designated by reference numeral 46 and outlined by a broken line, having a nucleotide sequence herein symbolically represented by A'.

[0297] The fibroblast genomic record 40 contains a binding site having a nucleotide sequence symbolically represented by A, which is complementary to the nucleotide sequence of A', and therefore the short RNA segment 46 binds to the fibroblast genomic record 40. This binding activates the fibroblast genomic record, causing cell 37 to differentiate into a fibroblast cell 3 (Fig. 1). Other genomic records, designated by reference numerals 39 and 41–44, comprise binding sites having nucleotide sequences that are symbolically represented by F, E, B, C and D respectively, which are not complementary to the nucleotide sequence of the short RNA segment 46 symbolically represented by A' and are therefore not activated by it. Genomic Records, such as the fibroblast genomic record 40, contain additional information, designated by reference numeral 45,

which is further elaborated hereinbelow with reference to Figs. 6 and 7.

[0298] Reference is now made to Fig. 5B, which is a simplified schematic diagram that illustrates a concept of cellular differentiation that is mediated by Genomic Records. Fig. 5B depicts two cells in an organism, cell A designated by reference numeral 47 and cell B designated by reference numeral 48, each having a genome 38. It is appreciated that since cell A 47 and cell B 48 are cells in the same organism, the genome 38 of cells 47 and 48 is identical. Despite having an identical genome 38, cell A 47 differentiates differently from cell B 48 due to the activation of different genomic records in these two cells. In cell A 47, the fibroblast genomic record 40 is activated, causing cell A 47 to differentiate into a fibroblast cell 3, whereas in cell B 48, the bone genomic record 42 is activated, causing cell B 48 to differentiate into a bone cell 4 (Fig. 1). The activation of different genomic records in these two cells is due to the different maternal short RNA segments which each received. Cell A 47 received a maternal short RNA segment designated 46 bearing a nucleotide sequence represented by A' that activates the fibroblast genomic record 40, whereas cell B 48 received a maternal short

RNA segment designated 49 bearing a nucleotide sequence represented by B' that activates the bone genomic record 42.

[0299] Reference is now made to Fig. 6 which is a schematic diagram illustrating a "genomically programmed cell differentiation" concept of the conceptual model of the present invention, addressing the genomic differentiation enigma.

[0300] A cell designated cell A 50 divides into 2 cells designated cell B 51 and cell C 52. Cell A 50, cell B 51 and cell C 52 each comprise a genome 38. Each genome 38 comprises a plurality of genomic records, herein exemplified by reference numerals 40, 42 and 43. It is appreciated that since cell A 50, cell B 51 and cell C 52 are cells in the same organism, the genome 38 of these cells, and the genomic records of these cells, exemplified by 40, 42 and 43, are identical.

[0301] As described above with reference to Fig. 5B, at its inception, cell A 50 receives a maternal short RNA segment, designated by reference numeral 46 and outlined by a broken line, having nucleotide sequence represented by A'. This short RNA sequence activates the fibroblast genomic record 40, thereby causing cell A 50 to differentiate into a fibroblast cell 3. However, Fig. 6 elaborates on

some of the additional information 45 of Fig. 5A of the genomic records. Specifically, a genomic record may also comprise two short genomic sequences, referred to here as Daughter Cell Genomic Addresses. Blocks designated B and C within the fibroblast genomic record in cell A 50 are Daughter Cell Genomic Addresses of the fibroblast genomic record. At cell division, each parent cell transcribes two short RNA segments, corresponding to the two Daughter Cell Genomic Addresses of the genomic record of that parent cell. The parent cell then transfers one of the Daughter Cell Genomic Addresses to each of its two daughter cells. As an example, cell A 50 transcribes and transfers to its two daughter cells 51 and 52 two short RNA segments, designated by reference numerals 49 and 53 and outlined by a broken line. The nucleotide sequences of these two short RNA segments, represented by B' and C' respectively, are complementary to the daughter cell genomic addresses designated B and C comprised in the fibroblast genomic record 40.

[0302] Cell B 51 therefore receives the abovementioned maternal short RNA segment designated 49, having a nucleotide sequence represented by B', which binds complementarily to the genomic address B of the bone genomic record 42.



The binding of the nucleotide sequence B' to the genomic address B activates this genomic record, which in turn causes cell B 51 to differentiate into a bone cell 4. Similarly, cell C 52 receives the abovementioned maternal short RNA segment designated 53 having a nucleotide sequence represented by C', which binds complementarily to the genomic address C of the cartilage genomic record 43. The binding of the nucleotide sequence C' to the genomic address C activates this genomic record, which in turn causes cell C 52 to differentiate into a cartilage cell 1 (Fig. 1).

[0303] It is appreciated that the mechanism illustrated by Fig. 6 enables the determination of the cell fate of an unlimited lineage of daughter cells containing the same DNA 5 (Fig. 1). For example, when cell B 51 and cell C 52 divide into their respective daughter cells (not shown), they transfer the short RNA segments designated by reference numerals 54–57 to their respective daughter cells. The genomic record that is activated in each of these daughter cells is affected by the identity of the maternal short RNA segments 54–57 that they each receive, which in turn determines their cell fate.

[0304] Reference is now made to Fig. 7 which is a schematic dia-

gram illustrating a "genomically programmed cell-specific protein expression modulation" concept of the conceptual model of the present invention, addressing the genomic differentiation enigma.

[0305] Cell A 58 receives a maternal short RNA segment designated 46 having a nucleotide sequence represented by A'. This maternal short RNA segment 46 activates the fibroblast genomic record 40 by complementarily binding to a binding site in the fibroblast genomic record, whose nucleotide sequence is designated A, and is complementary to the nucleotide sequence represented by A'. This is similar to the process shown in Fig. 5A. However, Fig. 7 further elaborates on some of the additional information 45 (Fig. 5A). The fibroblast genomic record 40 comprises three short nucleotide segments, whose nucleotide sequences are symbolically represented by 1, 2 and 4 respectively. These short nucleotide segments encode three respective short RNA oligonucleotides, designated by reference numerals 59–61. Each of these short RNA oligonucleotides modulates expression of a respective one of the target genes GENE 1, GENE 2 and GENE 4, designated by reference numerals 62–64 respectively, by complementarily binding to a binding site sequence associated with that

target gene. In a preferred embodiment of the present invention, the translation inhibition of target genes by complementarily binding to binding sites located in UTRs of the target genes modulates the expression of target genes such as 62–64. Cell A 58 thus differentiates into a fibroblast cell 3 (see also Fig. 1) because the expression of genes 1, 2 and 4 was modulated.

[0306] It is appreciated that the concept of genomic records is compatible with features of miRNA-like oligonucleotides of the present invention. A genomic record may comprise a cluster of short RNA segments that modulates the expression of target genes and thus influences differentiation. These features of genomic records are similar to the clusters of miRNA-like oligonucleotides of the present invention, which inhibit the translation of their respective target genes by complementarily binding to binding sites located in the of mRNA of these target genes.

[0307] Reference is now made to Fig. 8, which is a simplified diagram describing a plurality of novel bioinformatically-detected oligonucleotide of the present invention referred to here as the Genomic Address Messenger (GAM) oligonucleotide, which modulates the expression of respective target genes whose function and utility are known in the

art.

- [0308] GAM is a novel bioinformatically detectable regulatory, non-protein-coding, miRNA-like oligonucleotide. The method by which GAM is detected is described with additional reference to Figs. 8–15.
- [0309] The GAM PRECURSOR is encoded by the human genome. The GAM TARGET GENE is a gene encoded by the human genome.
- [0310] The GAM PRECURSOR encodes a GAM PRECURSOR RNA. Similar to other miRNA oligonucleotides, the GAM PRECURSOR RNA does not encode a protein.
- [0311] GAM PRECURSOR RNA folds onto itself, forming GAM FOLDED PRECURSOR RNA, which has a two-dimensional "hairpin" structure. GAM PRECURSOR RNA folds onto itself, forming GAM FOLDED PRECURSOR RNA, which has a two-dimensional "hairpin structure". As is well-known in the art, this "hairpin structure" is typical of RNA encoded by known miRNA precursor oligonucleotides and is due to the full or partial complementarity of the nucleotide sequence of the first half of an miRNA precursor to the RNA that is encoded by a miRNA oligonucleotide to the nucleotide sequence of the second half thereof.
- [0312] A complementary sequence is a sequence which is re-

versed and wherein each nucleotide is replaced by a complementary nucleotide, as is well known in the art (e.g. ATGGC is the complementary sequence of GCCAT).

[0313] An enzyme complex designated DICER COMPLEX, an enzyme complex composed of Dicer RNaseIII together with other necessary proteins, cuts the GAM FOLDED PRECURSOR RNA yielding a single-stranded ~22 nt-long RNA segment designated GAM RNA.

[0314] GAM TARGET GENE encodes a corresponding messenger RNA, designated GAM TARGET RNA. As is typical of mRNA of a protein-coding gene, each GAM TARGET RNAs of the present invention comprises three regions, as is typical of mRNA of a protein-coding gene: a 5' untranslated region, a protein-coding region and a 3' untranslated region, designated 5'UTR, PROTEIN-CODING and 3'UTR, respectively.

[0315] GAM RNA binds complementarily to one or more target binding sites located in the untranslated regions of each of the GAM TARGET RNAs of the present invention. This complementary binding is due to the partial or full complementarity between the nucleotide sequence of GAM RNA and the nucleotide sequence of each of the target binding sites. As an illustration, Fig. 8 shows three such

target binding sites, designated BINDING SITE I, BINDING SITE II and BINDING SITE III, respectively. It is appreciated that the number of target binding sites shown in Fig. 8 is only illustrative and that any suitable number of target binding sites may be present. It is further appreciated that although Fig. 8 shows target binding sites only in the 3'UTR region, these target binding sites may instead be located in the 5'UTR region or in both the 3'UTR and 5'UTR regions.

[0316] The complementary binding of GAM RNA to target binding sites on GAM TARGET RNA, such as BINDING SITE I, BINDING SITE II and BINDING SITE III, inhibits the translation of each of the GAM TARGET RNAs of the present invention into respective GAM TARGET PROTEIN, shown surrounded by a broken line.

[0317] It is appreciated that the GAM TARGET GENE in fact represents a plurality of GAM target genes. The mRNA of each one of this plurality of GAM target genes comprises one or more target binding sites, each having a nucleotide sequence which is at least partly complementary to GAM RNA and which when bound by GAM RNA causes inhibition of translation of the GAM target mRNA into a corresponding GAM target protein.

- [0318] The mechanism of the translational inhibition that is exerted by GAM RNA on one or more GAM TARGET GENES may be similar or identical to the known mechanism of translational inhibition exerted by known miRNA oligonucleotides.
- [0319] The nucleotide sequences of each of a plurality of GAM oligonucleotides that are described by Fig. 8 and their respective genomic sources and genomic locations are set forth in Tables 1-3, hereby incorporated herein.
- [0320] The nucleotide sequences of GAM PRECURSOR RNAs, and a schematic representation of a predicted secondary folding of GAM FOLDED PRECURSOR RNAs, of each of a plurality of GAM oligonucleotides that are described by Fig. 8 are set forth in Table 4, hereby incorporated herein.
- [0321] The nucleotide sequences of "diced" GAM RNAs of each of a plurality of GAM oligonucleotides that are described by Fig. 8 are set forth in Table 5, hereby incorporated herein.
- [0322] The nucleotide sequences of target binding sites, such as BINDING SITE I, BINDING SITE II and BINDING SITE III that are found on GAM TARGET RNAs of each of a plurality of GAM oligonucleotides that are described by Fig. 8, and a schematic representation of the complementarity of each of these target binding sites to each of a plurality of GAM

RNAs that are described by Fig. 8 are set forth in Tables 6–7, hereby incorporated herein.

[0323] It is appreciated that the specific functions and accordingly the utilities of each of a plurality of GAM oligonucleotides that are described by Fig. 8 are correlated with and may be deduced from the identity of the GAM TARGET GENES inhibited thereby, and whose functions are set forth in Table 8, hereby incorporated herein.

[0324] Studies documenting the well known correlations between each of a plurality of GAM TARGET GENES that are described by Fig. 8 and the known gene functions and related diseases are listed in Table 9, hereby incorporated herein.

[0325] The present invention discloses a novel group of human oligonucleotides, belonging to the miRNA-like oligonucleotide group, here termed GAM oligonucleotides, for which a specific complementary binding has been determined bioinformatically.

[0326] Reference is now made to Fig. 9 which is a simplified block diagram illustrating a bioinformatic oligonucleotide detection system and method constructed and operative in accordance with a preferred embodiment of the present invention.



- [0327] An important feature of the present invention is a bioinformatic oligonucleotide detection engine 100, which is capable of bioinformatically detecting oligonucleotides of the present invention.
- [0328] The functionality of the bioinformatic oligonucleotide detection engine 100 includes receiving expressed RNA data 102, sequenced DNA data 104, and PROTEIN FUNCTION DATA 106; performing a complex process of analysis of this data as elaborated hereinbelow, and based on this analysis provides information, designated by reference numeral 108, identifying and describing features of novel oligonucleotides.
- [0329] Expressed RNA data 102 comprises published expressed sequence tags (EST) data, published mRNA data, as well as other published RNA data. Sequenced DNA data 104 comprises alphanumeric data representing genomic sequences and preferably including annotations such as information indicating the location of known protein-coding regions relative to the genomic sequences.
- [0330] PROTEIN FUNCTION DATA 106 comprises information from scientific publications e.g. physiological functions of known proteins and their connection, involvement and possible utility in treatment and diagnosis of various dis-

eases.

[0331] Expressed RNA data 102 and sequenced DNA data 104 may preferably be obtained from data published by the National Center for Biotechnology Information (NCBI) at the National Institute of Health (NIH) (Jenuth J.P. (2000). *Methods Mol. Biol.* 132:301–312(2000), herein incorporated by reference) as well as from various other published data sources. PROTEIN FUNCTION DATA 106 may preferably be obtained from any one of numerous relevant published data sources, such as the Online Mendelian Inherited Disease In Man (OMIM(TM), Hamosh et al., *Nucleic Acids Res.* 30: 52–55(2002)) database developed by John Hopkins University, and also published by NCBI (2000).

[0332] Prior to or during actual detection of BIOINFORMATI-  
CALLY-DETECTED GROUP OF NOVEL OLIGONUCLEOTIDES  
108 by the bioinformatic oligonucleotide detection engine  
100, bioinformatic oligonucleotide detection engine training & validation functionality 110 is operative. This functionality uses one or more known miRNA oligonucleotides as a training set to train the bioinformatic oligonucleotide detection engine 100 to bioinformatically recognize miRNA-like oligonucleotides, and their respective potential target binding sites. BIOINFORMATIC OLIGONU-

CLEOTIDE DETECTION ENGINE TRAINING & VALIDATION  
FUNCTIONALITY 110 is further described herein below with  
reference to Fig. 10.

- [0333] The bioinformatic oligonucleotide detection engine 100 preferably comprises several modules which are preferably activated sequentially, and are described as follows:
- [0334] A NON-CODING GENOMIC SEQUENCE DETECTOR 112 operative to bioinformatically detect non-protein-coding genomic sequences. The non-protein-coding genomic sequence detector 112 is further described herein below with reference to Figs. 11A and 11B.
- [0335] A hairpin detector 114 operative to bioinformatically detect genomic "hairpin-shaped" sequences, similar to GAM FOLDED PRECURSOR RNA (Fig. 8). The hairpin detector 114 is further described herein below with reference to Figs. 12A and 12B.
- [0336] A Dicer-cut location detector 116 operative to bioinformatically detect the location on a GAM FOLDED PRECURSOR RNA which is enzymatically cut by DICER COMPLEX (Fig. 8), yielding "diced" GAM RNA. The Dicer-cut location detector 116 is further described herein below with reference to Figs. 13A-13C.
- [0337] A target gene binding site detector 118 operative to

bioinformatically detect target genes having binding sites, the nucleotide sequence of which is partially complementary to that of a given genomic sequence, such as a nucleotide sequence cut by DICER COMPLEX. The target gene binding site detector 118 is further described hereinbelow with reference to Figs. 14A and 14B.

[0338] A function & utility analyzer, designated by reference numeral 120, is operative to analyze the function and utility of target genes in order to identify target genes which have a significant clinical function and utility. The function & utility analyzer 120 is further described hereinbelow with reference to Fig. 15

[0339] According to an embodiment of the present invention, the bioinformatic oligonucleotide detection engine 100 may employ a cluster of 40 personal computers (PCs; XEON (R), 2.8GHz, with 80GB storage each) connected by Ethernet to eight servers (2-CPU, XEON (TM) 1.2-2.2GHz, with ~200GB storage each) and combined with an 8-processor server (8-CPU, Xeon 550Mhz w/ 8GB RAM) connected via 2 HBA fiber-channels to an EMC CLARIION (TM) 100-disks, 3.6 Terabyte storage device. A preferred embodiment of the present invention may also preferably comprise software that utilizes a commercial database

software program, such as MICROSOFT (TM) SQL Server 2000.

[0340] According to a preferred embodiment of the present invention, the bioinformatic oligonucleotide detection engine 100 may employ a cluster of 80 Servers (XEON (R), 2.8GHz, with 80GB storage each) connected by Ethernet to eight servers (2-CPU, XEON (TM) 1.2-2.2GHz, with ~200GB storage each) and combined with storage device (Promise Technology Inc., RM8000) connected to an 8-disks, 2 Terabytes total. A preferred embodiment of the present invention may also preferably comprise software that utilizes a commercial database software program, such as MICROSOFT (TM) SQL Server 2000. It is appreciated that the abovementioned hardware configuration is not meant to be limiting and is given as an illustration only. The present invention may be implemented in a wide variety of hardware and software configurations.

[0341] The present invention discloses 122764 novel oligonucleotides of the GAM group of oligonucleotides, which have been detected bioinformatically and 18602 novel polynucleotides of the GR group of polynucleotides, which have been detected bioinformatically. Laboratory confirmation of bioinformatically predicted oligonucleotides of

the GAM group of oligonucleotides, and several bioinformatically predicted polynucleotides of the GR group of polynucleotides, is described hereinbelow with reference to Figs. 21–24D, Fig. 27 and TABLE\_13.txt.

[0342] Reference is now made to Fig. 10 which is a simplified flowchart illustrating operation of a preferred embodiment of the bioinformatic oligonucleotide detection engine training & validation functionality 110 described hereinabove with reference to Fig. 9.

[0343] bioinformatic oligonucleotide detection engine training & validation functionality 110 begins by training the bioinformatic oligonucleotide detection engine 100 (Fig. 9) to recognize one or more known miRNA oligonucleotides, as designated by reference numeral 122. This training step comprises hairpin detector training & validation functionality 124, further described hereinbelow with reference to Fig. 12A, Dicer-cut location detector training & validation functionality 126, further described hereinbelow with reference to Fig. 13A and 13B, and target gene binding site detector training & validation functionality 128, further described hereinbelow with reference to Fig. 14A.

[0344] Next, the bioinformatic oligonucleotide detection engine training & validation functionality 110 is operative bioin-

formatically detect novel oligonucleotides, using bioinformatic oligonucleotide detection engine 100 (Fig. 9), as designated by reference numeral 130. Wet lab experiments are preferably conducted in order to validate expression and preferably function of some samples of the novel oligonucleotides detected by the bioinformatic oligonucleotide detection engine 100, as designated by reference numeral 132. Figs. 22A–24D and Table 13 illustrate examples of wet lab validation of sample novel human oligonucleotides bioinformatically-detected in accordance with a preferred embodiment of the present invention.

[0345] Reference is now made to Fig. 11A which is a simplified block diagram of a preferred implementation of the non-protein-coding genomic sequence detector 112 described hereinabove with reference to Fig. 9. The non-protein-coding genomic sequence detector 112 preferably receives at least two types of published genomic data: Expressed RNA data 102 and sequenced DNA data 104. The expressed RNA data 102 may include, inter alia, EST data, EST clusters data, EST genome alignment data and mRNA data. Sources for expressed RNA data 102 include NCBI dbEST, NCBI UniGene clusters and mapping data, and TIGR

gene indices (Kirkness F. and Kerlavage, A.R., *Methods Mol. Biol.* 69:261–268 (1997)). Sequenced DNA data 104 may include sequence data (FASTA format files), and feature annotations (GenBank file format) mainly from NCBI databases. Based on the above mentioned input data, the non-protein-coding genomic sequence detector 112 produces a plurality of non-protein-coding genomic sequences 136. Preferred operation of the non-protein-coding genomic sequence detector 112 is described hereinbelow with reference to Fig. 11B.

[0346] Reference is now made to Fig. 11B which is a simplified flowchart illustrating a preferred operation of the non-protein-coding genomic sequence detector 112 of Fig. 9. Detection of non-protein-coding genomic sequences 136, generally preferably progresses along one of the following two paths:

[0347] A first path for detecting non-protein-coding genomic sequences 136 (Fig. 11A) begins with receipt of a plurality of known RNA sequences, such as EST data. Each RNA sequence is first compared with known protein-coding DNA sequences, in order to select only those RNA sequences which are non-protein-coding, i.e. intergenic or intronic sequences. This can preferably be performed by using one



of many alignment algorithms known in the art, such as BLAST (Altschul et al., J. Mol. Biol. 215:403–410 (1990)). This sequence comparison preferably also provides localization of the RNA sequence on the DNA sequences.

[0348] Alternatively, selection of non-protein-coding RNA sequences and their localization on the DNA sequences can be performed by using publicly available EST cluster data and genomic mapping databases, such as the UNIGENE database published by NCBI or the TIGR database. Such databases, map expressed RNA sequences to DNA sequences encoding them, find the correct orientation of EST sequences, and indicate mapping of ESTs to protein-coding DNA regions, as is well known in the art. Public databases, such as TIGR, may also be used to map an EST to a cluster of ESTs, known in the art as Tentative Human Consensus and assumed to be expressed as one segment. Publicly available genome annotation databases, such as NCBI's GenBank, may also be used to deduce expressed intronic sequences.

[0349] Optionally, an attempt may be made to "expand" the non-protein RNA sequences thus found, by searching for transcription start and end signals, respectively upstream and downstream of the location of the RNA on the DNA, as is

well known in the art.

[0350] A second path for detecting non-protein-coding genomic sequences 136 (Fig. 11A) begins with receipt of DNA sequences. The DNA sequences are parsed into non-protein-coding sequences, using published DNA annotation data, by extracting those DNA sequences which are between known protein-coding sequences. Next, transcription start and end signals are sought. If such signals are found, and depending on their robustness, probable expressed non-protein-coding genomic sequences are obtained. Such approach is especially useful for identifying novel GAM oligonucleotides which are found in proximity to other known miRNA oligonucleotides, or other wet lab validated GAM oligonucleotides. Since, as described hereinbelow with reference to Fig. 16, GAM oligonucleotides are frequently found in clusters; sequences located near known miRNA oligonucleotides are more likely to contain novel GAM oligonucleotides. Optionally, sequence orthology, i.e. sequence conservation in an evolutionary related species, may be used to select genomic sequences having a relatively high probability of containing expressed novel GAM oligonucleotides.

[0351] Reference is now made to Fig. 12A which is a simplified

block diagram of a preferred implementation of the hairpin detector 114 described hereinabove with reference to Fig. 9.

[0352] The goal of the hairpin detector 114 is to detect hairpin-shaped genomic sequences, similar to those of known miRNA oligonucleotides. A hairpin-shaped genomic sequence is a genomic sequence, having a first half which is at least partially complementary to a second half thereof, which causes the halves to fold onto themselves, thereby forming a hairpin structure, as mentioned hereinabove with reference to Fig. 8.

[0353] The hairpin detector 114 (Fig. 9) receives a plurality of non-protein-coding genomic sequences 136 (Fig. 11A). Following operation of hairpin detector training & validation functionality 124 (Fig. 10), the hairpin detector 114 is operative to detect and output hairpin-shaped sequences, which are found in the non-protein-coding genomic sequences 136. The hairpin-shaped sequences detected by the hairpin detector 114 are designated hairpin structures on genomic sequences 138. A preferred mode of operation of the hairpin detector 114 is described hereinbelow with reference to Fig. 12B.

[0354] hairpin detector training & validation functionality 124 in-

cludes an iterative process of applying the hairpin detector 114 to known hairpin-shaped miRNA precursor sequences, calibrating the hairpin detector 114 such that it identifies a training set of known hairpin-shaped miRNA precursor sequences, as well as other similarly hairpin-shaped sequences. In a preferred embodiment of the present invention, the hairpin detector training & validation functionality 124 trains the hairpin detector 114 and validates each of the steps of operation thereof described hereinbelow with reference to Fig. 12B

[0355] The hairpin detector training & validation functionality 124 preferably uses two sets of data: the aforesaid training set of known hairpin-shaped miRNA precursor sequences, such as hairpin-shaped miRNA precursor sequences of 440 miRNA oligonucleotides of *H. sapiens*, *M. musculus*, *C. elegans*, *C. Brigssae* and *D. Melanogaster*, annotated in the RFAM database (Griffiths-Jones 2003), and a background set of about 1000 hairpin-shaped sequences found in expressed non-protein-coding human genomic sequences. The background set is expected to comprise some valid, previously undetected hairpin-shaped miRNA-like precursor sequences, and many hairpin-shaped sequences which are not hairpin-shaped

miRNA-like precursors.

[0356] In a preferred embodiment of the present invention the efficacy of the hairpin detector 114 (Fig. 9) is confirmed. For example, when a similarity threshold is chosen such that 87% of the known hairpin-shaped miRNA precursors are successfully predicted, only 21.8% of the 1000 background set of hairpin-shaped sequences are predicted to be hairpin-shaped miRNA-like precursors.

[0357] Reference is now made to Fig. 12B which is a simplified flowchart illustrating preferred operation of the hairpin detector 114 of Fig. 9. The hairpin detector 114 preferably initially uses a secondary structure folding algorithm based on free-energy minimization, such as the MFOLD algorithm, described in Mathews et al. J. Mol. Biol. 288:911-940 (1999) and Zuker, M. Nucleic Acids Res. 31: 3406-3415 (2003), the disclosure of which is hereby incorporated by reference. This algorithm is operative to calculate probable secondary structure folding patterns of the non-protein-coding genomic sequences 136 (Fig. 11A) as well as the free-energy of each of these probable secondary folding patterns. The secondary structure folding algorithm, such as the MFOLD algorithm (Mathews, 1997; Zuker 2003), typically provides a listing of the

base-pairing of the folded shape, i.e. a listing of each pair of connected nucleotides in the sequence.

[0358] Next, the hairpin detector 114 analyzes the results of the secondary structure folding patterns, in order to determine the presence and location of hairpin folding structures. The goal of this second step is to assess the base-pairing listing provided by the secondary structure folding algorithm, in order to determine whether the base-pairing listing describes one or more hairpin type bonding pattern. Preferably, sequence segment corresponding to a hairpin structure is then separately analyzed by the secondary structure folding algorithm in order to determine its exact folding pattern and free-energy.

[0359] The hairpin detector 114 then assesses the hairpin structures found by the previous step, comparing them to hairpin structures of known miRNA precursors, using various characteristic hairpin structure features such as its free-energy and its thermodynamic stability, the amount and type of mismatched nucleotides and the existence of sequence repeat-elements, number of mismatched nucleotides in positions 18-22 counting from loop, and Percent of G nucleotide. Only hairpins that bear statistically significant resemblance to the training set of hairpin

structures of known miRNA precursors, according to the abovementioned parameters, are accepted.

[0360] In a preferred embodiment of the present invention, similarity to the training set of hairpin structures of known miRNA precursors is determined using a "similarity score" which is calculated using a multiplicity of terms, where each term is a function of one of the abovementioned hairpin structure features. The parameters of each function are found heuristically from the set of hairpin structures of known miRNA precursors, as described hereinabove with reference to hairpin detector training & validation functionality 124 (Fig. 10). The selection of the features and their function parameters is optimized so as to achieve maximized separation between the distribution of similarity scores validated miRNA precursor hairpin structures, and the distribution of similarity scores of hairpin structures detected in the background set mentioned hereinabove with reference to Fig. 12B.

[0361] In an alternative preferred embodiment of the present invention, the step described in the preceding paragraph may be split into two stages. A first stage implements a simplified scoring method, typically based on thresholding a subset of the hairpin structure features described

hereinabove, and may employ a minimum threshold for hairpin structure length and a maximum threshold for free energy. A second stage is preferably more stringent, and preferably employs a full calculation of the weighted sum of terms described hereinabove. The second stage preferably is performed only on the subset of hairpin structures that survived the first stage.

[0362] The hairpin detector 114 also attempts to select hairpin structures whose thermodynamic stability is similar to that of hairpin structures of known miRNA precursors. This may be achieved in various ways. A preferred embodiment of the present invention utilizes the following methodology, preferably comprising three logical steps:

[0363] First, the hairpin detector 114 attempts to group hairpin structures into "families" of closely related hairpin structures. As is known in the art, a secondary structure folding algorithm typically provides multiple alternative folding patterns, for a given genomic sequence and indicates the free energy of each alternative folding pattern. It is a particular feature of the present invention that the hairpin detector 114 preferably assesses the various hairpin structures appearing in the various alternative folding patterns and groups' hairpin structures which appear at



identical or similar sequence locations in various alternative folding patterns into common sequence location based "families" of hairpins. For example, all hairpin structures whose center is within 7 nucleotides of each other may be grouped into a "family". Hairpin structures may also be grouped into a "family" if their nucleotide sequences are identical or overlap to a predetermined degree.

[0364] It is also a particular feature of the present invention that the hairpin structure "families" are assessed in order to select only those families which represent hairpin structures that are as thermodynamically stable as those of hairpin structures of known miRNA precursors. Preferably only families which are represented in at least a selected majority of the alternative secondary structure folding patterns, typically 65%, 80% or 100% are considered to be sufficiently stable. Our tests suggest that only about 50% of the hairpin structures, predicted by the MFOLD algorithm with default parameters, are members of sufficiently stable families, comparing to about 90% of the hairpin structures that contain known miRNAs. This percent depends on the size of the fraction that was fold. In an alternative embodiment of the present invention we use frac-

tions of size 1000 nts as preferable size. Different embodiment uses other sizes of genomics sequences, more or less strict demand for representation in the alternative secondary structure folding patterns.

[0365] It is an additional particular feature of the present invention that the most suitable hairpin structure is selected from each selected family. For example, a hairpin structure which has the greatest similarity to the hairpin structures appearing in alternative folding patterns of the family may be preferred. Alternatively or additionally, the hairpin structures having relatively low free energy may be preferred.

[0366] Alternatively or additionally considerations of homology to hairpin structures of other organisms and the existence of clusters of thermodynamically stable hairpin structures located adjacent to each other along a sequence may be important in selection of hairpin structures. The tightness of the clusters in terms of their location and the occurrence of both homology and clusters may be of significance.

[0367] Reference is now made to Figs. 13A–13C, which together describe the structure and operation of the Dicer-cut location detector 116, described hereinabove with reference

to Fig. 9.

[0368] Reference is now made to Fig. 13A, which is a simplified block diagram of a preferred implementation of the Dicer-cut location detector 116. The goal of the Dicer-cut location detector 116 is to detect the location in which the DICER COMPLEX, described hereinabove with reference to Fig. 8, dices GAM FOLDED PRECURSOR RNA, yielding GAM RNA.

[0369] The Dicer-cut location detector 116 therefore receives a plurality of hairpin structures on genomic sequences, designated by reference numeral 138 (Fig. 12A), and following operation of Dicer-cut location detector training & validation functionality 126 (Fig 10), is operative to detect a plurality of Dicer-cut sequences from hairpin structures, designated by reference numeral 140.

[0370] Reference is now made to Fig. 13B, which is a simplified flowchart illustrating a preferred implementation of Dicer-cut location detector training & validation functionality 126.

[0371] A general goal of the Dicer-cut location detector training & validation functionality 126 is to analyze the Dicer-cut locations of known diced miRNA on respective hairpin-shaped miRNA precursors in order to determine a com-

mon pattern in these locations, which can be used to predict Dicer-cut locations on GAM folded precursor RNAs.

[0372] The Dicer-cut locations of known miRNA precursors are obtained and studied. Locations of the 5' and/or 3' ends of the known diced miRNA oligonucleotides are preferably represented by their respective distances from the 5' end of the corresponding hairpin-shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNA oligonucleotides are preferably represented by the relationship between their locations and the locations of one or more nucleotides along the hairpin-shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNA oligonucleotides are preferably represented by the relationship between their locations and the locations of one or more bound nucleotide pairs along the hairpin-shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNA oligonucleotides are preferably represented by the relationship between their locations and the locations of one or more mismatched nucleotide pairs along the hairpin-shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNA oligonucleotides are preferably

represented by the relationship between their locations and the locations of one or more unmatched nucleotides along the hairpin-shaped miRNA precursor. Additionally or alternatively, locations of the 5' and/or 3' ends of the known diced miRNA oligonucleotides are preferably represented by their respective distances from the loop located at the center of the corresponding hairpin-shaped miRNA precursor.

[0373] One or more of the foregoing location metrics may be employed in the Dicer-cut location detector training & validation functionality 126. Additionally, metrics related to the nucleotide content of the diced miRNA and/or of the hairpin-shaped miRNA precursor may be employed.

[0374] In a preferred embodiment of the present invention, Dicer-cut location detector training & validation functionality 126 preferably employs standard machine learning techniques known in the art of machine learning to analyze existing patterns in a given "training set" of examples. Standard machine learning techniques are capable, to a certain degree, of detecting patterns in examples to which they have not been previously exposed that are similar to those in the training set. Such machine learning techniques include, but are not limited to neural net-

works, Bayesian Modeling, Bayesian Networks, Support Vector Machines (SVM), Genetic Algorithms, Markovian Modeling, Maximum Likelihood Modeling, Nearest Neighbor Algorithms, Decision Trees and other techniques, as is well-known in the art.

[0375] In accordance with an embodiment of the present invention, two or more classifiers or predictors based on the abovementioned machine learning techniques are separately trained on the abovementioned training set, and are used jointly in order to predict the Dicer-cut location. As an example, Fig. 13B illustrates operation of two classifiers, a 3' end recognition classifier and a 5' end recognition classifier. Most preferably, the Dicer-cut location detector training & validation functionality 126 implements a "best-of-breed" approach employing a pair of classifiers based on the abovementioned Bayesian Modeling and Nearest Neighbor Algorithms, and accepting only "potential GAM RNAs" that score highly on one of these predictors. In this context, "high scores" means scores that have been demonstrated to have low false positive value when scoring known miRNA oligonucleotides. Alternatively, the Dicer-cut location detector training & validation functionality 126 may implement operation of more or less than

two classifiers.

[0376] Predictors used in a preferred embodiment of the present invention are further described hereinbelow with reference to Fig. 13C. A computer program listing of a computer program implementation of the Dicer-cut location detector training & validation functionality 126 is enclosed on an electronic medium in computer-readable form, and is hereby incorporated by reference herein.

[0377] When evaluated on the abovementioned validation set of 440 published miRNA oligonucleotides using k-fold cross validation (Mitchell, 1997) with  $k = 3$ , the performance of the resulting predictors is as follows: In 70% of known miRNA oligonucleotides, a 5' end location is correctly determined by a Support Vector Machine predictor within up to two nucleotides; a Nearest Neighbor (EDIT DISTANCE) predictor achieves 56% accuracy (247/440); and a Two-Phased Predictor that uses Bayesian modeling (TWO PHASED) achieves 80% accuracy (352/440) when only the first phase is used. When the second phase (strand choice) is implemented by a naive Bayesian model, the accuracy is 55% (244/440), and when the K-nearest-neighbor modeling is used for the second phase, 374/440 decisions are made and the accuracy is 65% (242/374). A K-near-

est-neighbor predictor (FIRST-K) achieves 61% accuracy (268/440). The accuracies of all predictors are considerably higher on top-scoring subsets of published miRNA oligonucleotides.

[0378] Finally, in order to validate the efficacy and accuracy of the Dicer-cut location detector 116, a sample of novel oligonucleotides detected thereby is preferably selected, and validated by wet lab experiments. Laboratory results validating the efficacy of the Dicer-cut location detector 116 are described hereinbelow with reference to Figs. 22-24D, Fig27 and also in the enclosed file "TA-BLE\_13.txt".

[0379] Reference is now made to Fig. 13C, which is a simplified flowchart illustrating an operation of a Dicer-cut location detector 116 (Fig. 9), constructed and operative in accordance with a preferred embodiment of the present invention. The Dicer-cut location detector 116 preferably comprises a machine learning computer program module, which is trained to recognize Dicer-cut locations on known hairpin-shaped miRNA precursors, and based on this training, is operable to detect Dicer-cut locations of novel GAM RNA (Fig. 8) on GAM FOLDED PRECURSOR RNA (Fig. 8). In a preferred embodiment of the present inven-



tion, the Dicer-cut location module preferably utilizes machine learning algorithms, including but not limited to Support Vector Machine, Bayesian modeling, Nearest Neighbors, and K-nearest-neighbor algorithms that are known in the art.

[0380] When initially assessing a novel GAM FOLDED PRECURSOR RNA, each 19–24 nt-long segment thereof is considered to be a potential GAM RNA, because the Dicer-cut location is initially unknown.

[0381] For each such potential GAM RNA, the location of its 5' end or the locations of its 5' and 3' ends are scored by at least one recognition classifier or predictor, operating on features such as the following: Locations of the 5' and/or 3' ends of the known diced miRNA oligonucleotides, which are preferably represented by their respective distances from the 5' end of the corresponding hairpin-shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNA oligonucleotides, which are preferably represented by the relationship between their locations and the locations of one or more nucleotides along the hairpin-shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNA oligonucleotides, which are prefer-

ably represented by the relationship between their locations and the locations of one or more bound nucleotide pairs along the hairpin-shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNA oligonucleotides, which are preferably represented by the relationship between their locations and the locations of one or more mismatched nucleotide pairs along the hairpin-shaped miRNA precursor. Additionally or alternatively, the 5' and/or 3' ends of the known diced miRNA oligonucleotides, which are preferably represented by the relationship between their locations and the locations of one or more unmatched nucleotides along the hairpin-shaped miRNA precursor. Additionally or alternatively, locations of the 5' and/or 3' ends of the known diced miRNA oligonucleotides, which are preferably represented by their respective distances from the loop located at the center of the corresponding hairpin-shaped miRNA precursor. Additionally or alternatively, metrics related to the nucleotide content of the diced miRNA and/or of the hairpin-shaped miRNA precursor.

[0382] In a preferred embodiment of the present invention, the Dicer-cut location detector 116 (Fig. 9) may use a Support

Vector Machine predictor.

[0383] In another preferred embodiment of the present invention, the Dicer-cut location detector 116 (Fig. 9) preferably employs an "EDIT DISTANCE" predictor, which seeks sequences that are similar to those of known miRNA oligonucleotides, utilizing a Nearest Neighbor algorithm, where a similarity metric between two sequences is a variant of the Edit Distance algorithm (Gusfield, 1997). The EDIT DISTANCE predictor is based on an observation that miRNA oligonucleotides tend to form clusters, the members of which show marked sequence similarity.

[0384] In yet another preferred embodiment of the present invention, the Dicer-cut location detector 116 (Fig. 9) preferably uses a "TWO PHASE" predictor, which predicts the Dicer-cut location in two distinct phases: (a) selecting a double-stranded segment of the GAM FOLDED PRECURSOR RNA (Fig. 8) comprising the GAM RNA by naive Bayesian modeling and (b) detecting which strand of the double-stranded segment contains GAM RNA (Fig. 8) by employing either naive or K-nearest-neighbor modeling. K-nearest-neighbor modeling is a variant of the "FIRST-K" predictor described hereinbelow, with parameters optimized for this specific task. The "TWO PHASE" predictor

may be operated in two modes: either utilizing only the first phase and thereby producing two alternative Dicer-cut location predictions, or utilizing both phases and thereby producing only one final Dicer-cut location.

[0385] In still another preferred embodiment of the present invention, the Dicer-cut location detector 116 preferably uses a "FIRST-K" predictor, which utilizes a K-nearest-neighbor algorithm. The similarity metric between any two sequences is  $1 - E/L$ , where L is a parameter, preferably 8-10 and E is the edit distance between the two sequences, taking into account only the first L nucleotides of each sequence. If the K-nearest-neighbor scores of two or more locations on the GAM FOLDED PRECURSOR RNA (Fig. 8) are not significantly different, these locations are further ranked by a Bayesian model, similar to the one described hereinabove.

[0386] In accordance with an embodiment of the present invention, scores of two or more of the abovementioned classifiers or predictors are integrated, yielding an integrated score for each potential GAM RNA. As an example, Fig. 13C illustrates an integration of scores from two classifiers, a 3' end recognition classifier and a 5' end recognition classifier, the scores of which are integrated to yield

an integrated score. Most preferably, the INTEGRATED SCORE of Fig. 13C preferably implements a "best-of-breed" approach employing a pair of classifiers and accepting only "potential GAM RNAs" that score highly on one of the abovementioned "EDIT DISTANCE", or "TWO PHASE" predictors. In this context, "high scores" means scores that have been demonstrated to have low false positive value when scoring known miRNA oligonucleotides. Alternatively, the INTEGRATED SCORE may be derived from operation of more or less than two classifiers.

[0387] The INTEGRATED SCORE is evaluated as follows: (a) the "potential GAM RNA" having the highest score is preferably taken to be the most probable GAM RNA, and (b) if the integrated score of this most probable GAM RNA is higher than a pre-defined threshold, then the most probable GAM RNA is accepted as a PREDICTED GAM RNA. Preferably, this evaluation technique is not limited to the highest scoring potential GAM RNA.

[0388] In a preferred embodiment of the present invention, PREDICTED GAM RNAs comprising a low complexity nucleotide sequence (e.g., ATATATA) may optionally be filtered out, because there is a high probability that they are

part of a repeated element in the DNA, and are therefore not functional, as is known in the art. For each PREDICTED GAM RNA sequence, the number of occurrences of each two nt combination (AA, AT, AC) comprised in that sequence is counted. PREDICTED GAM RNA sequences where the sum of the two most probable combinations is higher than a threshold, preferably 8–10, are filtered out. As an example, when the threshold is set such that 2% of the known miRNA oligonucleotides are filtered out, 30% of the predicted GAM RNAs are filtered out.

[0389] Reference is now made to Fig. 14A, which is a simplified block diagram of a preferred implementation of the target gene binding site detector 118 described hereinabove with reference to Fig. 9. The goal of the target gene binding site detector 118 is to detect one or more binding sites located in 3'UTRs of the mRNA of a known gene, such as BINDING SITE I, BINDING SITE II and BINDING SITE III (Fig. 8), the nucleotide sequence of which binding sites is partially or fully complementary to a GAM RNA, thereby determining that the abovementioned known gene is a target gene of the GAM RNA.

[0390] The target gene binding site detector 118 (Fig. 9) receives a plurality of Dicer-cut sequences from hairpin structures

140 (Fig. 13A) and a plurality of potential target gene sequences 142, which are derived from sequenced DNA data 104 (Fig. 9).

[0391] The target gene binding site detector training & validation functionality 128 (Fig. 10) is operative to train the target gene binding site detector 118 on known miRNA oligonucleotides and their respective target genes and to build a background model for an evaluation of the probability of achieving similar results randomly (P value) for the target gene binding site detector 118 results. The target gene binding site detector training & validation functionality 128 constructs the model by analyzing both heuristically and computationally the results of the target gene binding site detector 118.

[0392] Following operation of target gene binding site detector training & validation functionality 128 (Fig. 10), the target gene binding site detector 118 is operative to detect a plurality of potential novel target genes having binding site/s 144, the nucleotide sequence of which is partially or fully complementary to that of each of the plurality of Dicer-cut sequences from hairpin structures 140. Preferred operation of the target gene binding site detector 118 is further described hereinbelow with reference to

Fig. 14B.

- [0393] Reference is now made to Fig. 14B, which is a simplified flowchart illustrating a preferred operation of the target gene binding site detector 118 of Fig. 9.
- [0394] In an embodiment of the present invention, the target gene binding site detector 118 first compares nucleotide sequences of each of the plurality of Dicer-cut sequences from hairpin structures 140 (Fig. 13A) to the potential target gene sequences 142 (Fig. 14A), such as 3' side UTRs of known mRNAs, in order to find crude potential matches. This step may be performed using a simple alignment algorithm such as BLAST.
- [0395] Then, the target gene binding site detector 118 filters these crude potential matches, to find closer matches, which more closely resemble published miRNA oligonucleotide binding sites.
- [0396] Next, the target gene binding site detector 118 expands the nucleotide sequences of the 3'UTR binding site found by the sequence comparison algorithm (e.g. BLAST or EDIT DISTANCE). A determination is made whether any subsequence of the expanded sequence may improve the match. The best match is considered the alignment.
- [0397] Free-energy and spatial structure are computed for the



resulting binding sites. Calculation of spatial structure may be performed by a secondary structure folding algorithm based on free-energy minimization, such as the MFOLD algorithm described in Mathews et al. (J. Mol. Biol. 288: 911-940 (1999)) and Zuker (Nucleic Acids Res. 31: 3406-3415 (2003)), the disclosure of which is hereby incorporated by reference. Free energy, spatial structure and the above preferences are reflected in scoring. The resulting scores are compared with scores characteristic of known binding sites of published miRNA oligonucleotides, and each binding site is given a score that reflects its resemblance to these known binding sites.

[0398] Finally, the target gene binding site detector 118 analyzes the spatial structure of the binding site. Each 3'UTR-GAM oligonucleotide pair is given a score. Multiple binding sites of the same GAM oligonucleotides to a 3'UTR are given higher scores than those that bind only once to a 3'UTR.

[0399] In a preferred embodiment of the present invention, performance of the target gene binding site detector 118 may be improved by integrating several of the abovementioned logical steps, using the methodology described hereinbelow.

[0400] For each of the dicer-cut sequence from hairpin structures 140, its starting segment, e.g. a segment comprising the first 8 nts from its 5' end, is obtained. For each starting segment, all of the 9 nt segments that are highly complementary to the starting segment are calculated. These calculated segments are referred to here as "potential binding site end segments". In a preferred embodiment of the present invention, for each 8 nt starting segment, the potential binding site end segments are all 9 nt segments whose complementary sequence contains a 7-9 nt sub-sequence that is not different from the starting segment by more than an insertion, deletion or replacement of one nt. Calculation of potential binding site end segments is preferably performed by a pre-processing tool that maps all possible 8 nt segments to their respective 9 nt segments.

[0401] Next, the mRNAs 3'UTRs is parsed into all the segments, with the same length as the potential binding site end segments, preferably 9 nt segments, comprised in the 3'UTR. Location of each such segment is noted, stored in a performance-efficient data structure and compared to the potential binding site end segments calculated in the previous step.

[0402] The target gene binding site detector 118 then expands the binding site sequence, preferably in the binding site 5' direction (i.e. immediately upstream), assessing the degree of its alignment to the dicer-cut sequence from hair-pin structures 140. Preferably, an alignment algorithm is implemented which uses specific weighting parameters based on an analysis of known miRNA oligonucleotide binding sites. As an example, it is apparent that a good match of the 3' end of the binding site is critically important, a match of the 5' end is less important but can compensate for a small number of mismatches at the 3' end of the binding site, and a match of the middle portion of the binding site is much less important.

[0403] Next, the number of binding sites found in a specific 3'UTR, the degree of alignment of each of these binding sites, and their proximity to each other are assessed and compared to these properties found in known binding sites of published miRNA oligonucleotides. In a preferred embodiment, the fact that many of the known binding sites are clustered is used to evaluate the P value of obtaining a cluster of a few binding sites on the same target gene 3'UTR in the following way. It scans different score thresholds and calculates for each threshold the number

and positions of possible binding sites with a score above the threshold. It then gets a P value for each threshold from a preprocessed calculated background matrix, described hereinbelow, and a number and positions of binding sites combination. The output score for each Dicer-cut sequences from hairpin structures 140 and potential target gene sequences 142 is the minimal P value, normalized with the number of threshold trails using a Bernoulli distribution. A preference of low P value pairs is made.

[0404] As mentioned hereinabove, for each target gene, a preprocessed calculated background matrix is built. The matrix includes rows for each number of miRNA oligonucleotide binding sites (in the preferred embodiment, the matrix includes 7 rows to accommodate 0 to 6 binding sites), and columns for each different score threshold (in the preferred embodiment, the matrix includes 5 columns for 5 different thresholds). Each matrix cell, corresponding to a specific number of binding sites and thresholds, is set to be the probability of getting equal or higher number binding sites and an equal or higher score using random 22 nt-long sequences with the same nucleotide distribution as known miRNA oligonucleotides (29.5% T,

24.5% A, 25% G and 21% C). Those probabilities are calculated by running the above procedure for 10000 random sequences that preserved the known miRNA nucleotide distribution (these sequence will be also referred to as miRNA oligonucleotide random sequences). The P value can be estimated as the number of random sequences that obeys the matrix cell requirement divided by the total number of random sequences (10000). In the preferred embodiment, 2 matrices are calculated. The P values of the second matrix are calculated under a constraint that at least two of the binding site positions are under a heuristically-determined constant value. The values of the second matrix are calculated without this constraint. The target gene binding site detector 118 uses the second matrix if the binding site positions agree with the constraint. Otherwise, it uses the first. In an alternative embodiment, only one matrix is calculated without any constraint on the binding sites positions.

[0405] A test performed using the target gene binding site detector 118 shows that all of the known miRNA oligonucleotide target genes are found using this algorithm with a P value of less than 0.5%. Running known miRNA oligonucleotides against 3400 potential 3'UTR of target

gene sequences yields on average 32 target genes for each miRNA oligonucleotide with a P value less than 0.5%, while background sequences, as well as inverse or complement sequence of known miRNA oligonucleotide (which preserve their high order sequence statistics) found, as expected, 17 target genes on average. This result reflects that the algorithm has the ability to detect real target genes with 47% accuracy.

[0406] Finally, orthology data may optionally be used to further prefer binding sites based on their conservation. Preferably, this may be used in cases such as (a) where both the target mRNA and miRNA oligonucleotide have orthologues in another organism, e.g. Human-Mouse orthology, or (b) where a miRNA oligonucleotide (e.g. viral miRNA oligonucleotide) targets two mRNAs in orthologous organisms. In such cases, binding sites that are conserved are preferred.

[0407] In accordance with another preferred embodiment of the present invention, binding sites may be searched by a reverse process. Sequences of K (preferably 22) nucleotides in a UTR of a target gene are assessed as potential binding sites. A sequence comparison algorithm, such as BLAST or EDIT DISTANCE variant, is then used to search elsewhere in the genome for partially or fully complemen-

tary sequences that are found in known miRNA oligonucleotides or computationally-predicted GAM oligonucleotides. Only complementary sequences that meet predetermined spatial structure and free-energy criteria as described hereinabove, are accepted. Clustered binding sites are strongly preferred and potential binding sites and potential GAM oligonucleotides that occur in evolutionarily-conserved genomic sequences are also preferred. Scoring of candidate binding sites takes into account free-energy and spatial structure of the binding site complexes, as well as the aforesaid preferences.

[0408] Reference is now made to Fig. 15 which is a simplified flowchart illustrating a preferred operation of the function & utility analyzer 120 described hereinabove with reference to Fig. 9. The goal of the function & utility analyzer 120 is to determine if a potential target gene is in fact a valid clinically useful target gene. Since a potential novel GAM oligonucleotide binding a binding site in the UTR of a target gene is understood to inhibit expression of that target gene, and if that target gene is shown to have a valid clinical utility, then in such a case it follows that the potential novel oligonucleotide itself also has a valid useful function which is the opposite of that of the target

gene.

[0409] The function & utility analyzer 120 preferably receives as input a plurality of potential novel target genes having binding site/s 144 (Fig. 14A), generated by the target gene binding site detector 118 (Fig. 9). Each potential oligonucleotide is evaluated as follows: First, the system checks to see if the function of the potential target gene is scientifically well established. Preferably, this can be achieved bioinformatically by searching various published data sources presenting information on known function of proteins. Many such data sources exist and are published as is well known in the art. Next, for those target genes the function of which is scientifically known and is well documented, the system then checks if scientific research data exists which links them to known diseases. For example, a preferred embodiment of the present invention utilizes the OMIM(TM) (Hamosh et al, 2002) database published by NCBI, which summarizes research publications relating to genes which have been shown to be associated with diseases. Finally, the specific possible utility of the target gene is evaluated. While this process too may be facilitated by bioinformatic means, it might require manual evaluation of published scientific research regard-



ing the target gene, in order to determine the utility of the target gene to the diagnosis and or treatment of specific disease. Only potential novel oligonucleotides, the target genes of which have passed all three examinations, are accepted as novel oligonucleotide.

[0410] Reference is now made to Fig. 16, which is a simplified diagram describing each of a plurality of novel bioinformatically-detected regulatory polynucleotide referred to in this Table as the Genomic Record (GR) polynucleotide. GR encodes an operon-like cluster of novel miRNA-like oligonucleotides, each of which in turn modulates the expression of at least one target gene. The function and utility of at least one target gene is known in the art.

[0411] The GR PRECURSOR is a novel, bioinformatically-detected, regulatory, non-protein-coding polynucleotide. The method by which the GR PRECURSOR is detected is described hereinabove with additional reference to Figs. 9-18.

[0412] The GR PRECURSOR encodes GR PRECURSOR RNA that is typically several hundred to several thousand nts long. The GR PRECURSOR RNA folds spatially, forming the GR FOLDED PRECURSOR RNA. It is appreciated that the GR FOLDED PRECURSOR RNA comprises a plurality of what is

known in the art as hairpin structures. Hairpin structures result from the presence of segments of the nucleotide sequence of GR PRECURSOR RNA in which the first half of each such segment has a nucleotide sequence which is at least a partial, and sometimes an accurate, reverse-complement sequence of the second half thereof, as is well known in the art.

[0413] The GR FOLDED PRECURSOR RNA is naturally processed by cellular enzymatic activity into a plurality of separate GAM precursor RNAs herein schematically represented by GAM1 FOLDED PRECURSOR RNA through GAM3 FOLDED PRECURSOR RNA. Each GAM folded precursor RNA is a hairpin-shaped RNA segment, corresponding to GAM FOLDED PRECURSOR RNA of Fig. 8.

[0414] The abovementioned GAM folded precursor RNAs are diced by DICER COMPLEX of Fig. 8, yielding short RNA segments of about 22 nts in length schematically represented by GAM1 RNA through GAM3 RNA. Each GAM RNA corresponds to GAM RNA of Fig. 8. GAM1 RNA, GAM2 RNA and GAM3 RNA each bind complementarily to binding sites located in the untranslated regions of their respective target genes, designated GAM1 TARGET RNA, GAM2 TARGET RNA and GAM3 TARGET RNA, respectively. These

target binding sites correspond to BINDING SITE I, BINDING SITE II and BINDING SITE III of Fig. 8. The binding of each GAM RNA to its target RNA inhibits the translation of its respective target proteins, designated GAM1 TARGET PROTEIN, GAM2 TARGET PROTEIN and GAM3 TARGET PROTEIN, respectively.

[0415] It is appreciated that the specific functions, and accordingly the utilities, of the GR polynucleotide are correlated with and may be deduced from the identity of the target genes that are inhibited by GAM RNAs that are present in the operon-like cluster of the polynucleotide. Thus, for the GR polynucleotide, schematically represented by GAM1 TARGET PROTEIN through GAM3 TARGET PROTEIN that are inhibited by the GAM RNA. The function of these target genes is elaborated in Table 8, hereby incorporated herein.

[0416] Reference is now made to Fig. 17 which is a simplified diagram illustrating a mode by which oligonucleotides of a novel group of operon-like polynucleotide described hereinabove with reference to Fig. 16 of the present invention, modulate expression of other such polynucleotide, in a cascading manner. GR1 PRECURSOR and GR2 PRECURSOR are two polynucleotides of the novel group of

operon-like polynucleotides designated GR PRECURSOR (Fig. 16). As is typical of polynucleotides of the GR group of polynucleotides GR1 PRECURSOR and GR2 PRECURSOR, each encode a long RNA precursor, which in turn folds into a folded RNA precursor comprising multiple hairpin shapes, and is cut into respective separate hairpin-shaped RNA segments, each of which RNA segments being diced to yield an oligonucleotide of a group of oligonucleotides designated GAM RNA. In this manner GR1 yields GAM1 RNA, GAM2 RNA and GAM3 RNA, and GR2 yields GAM4 RNA, GAM5 RNA and GAM6 RNA. As Fig. 17 shows, GAM3 RNA, which derives from GR1, binds a binding site located adjacent to GR2 GPRECURSOR thus modulating expression of GR2, thereby invoking expression of GAM4 RNA, GAM5 RNA and GAM6 RNA which derive from GR2. It is appreciated that the mode of modulation of expression presented by Fig. 17 enables an unlimited "cascading effect" in which a GR polynucleotide comprises multiple GAM oligonucleotides each of which may modulate expression of other GR polynucleotides each such GR polynucleotides comprising additional GAM oligonucleotide etc., whereby eventually certain GAM oligonucleotides modulate expression of target proteins.

[0417] This mechanism is in accord with the conceptual model of the present invention addressing the differentiation enigma, described hereinabove with specific reference to Figs. 6–7.

[0418] Reference is now made to Fig. 18 which is a block diagram illustrating an overview of a methodology for finding novel oligonucleotides and operon-like polynucleotides of the present invention, and their respective functions. According to a preferred embodiment of the present invention, the methodology to finding novel oligonucleotides of the present invention and their function comprises of the following major steps: First, FIND GAM OLIGONUCLEOTIDES 146 is used to detect, oligonucleotide of the novel group of oligonucleotide of the present invention, referred to here as GAM oligonucleotide. GAM oligonucleotides are located and their function elicited by detecting target proteins they bind and the function of those target proteins, as described hereinabove with reference to Figs. 9–15. Next, FIND GR POLYNUCLEOTIDES 147 is used to detect polynucleotide of a novel group of operon-like polynucleotide of the present invention, referred to here as GR polynucleotide. GR polynucleotides are located, by locating clusters of proximally located GAM oligonucleotide,

based on the previous step. Consequently, FIND HIERARCHY OF GR POLYNUCLEOTIDES 148 elicits the hierarchy of GR and GAM: binding sites for non-protein-binding GAM oligonucleotide comprised in each GR polynucleotide found are sought adjacent to other GR polynucleotides. When found, such a binding site indicates that the connection between the GAM and the GR the expression of which it modulates, and thus the hierarchy of the GR polynucleotides and the GAM oligonucleotides they comprise. Lastly, DEDUCE FUNCTION OF "HIGH" GR POLYNUCLEOTIDES AND GAM OLIGONUCLEOTIDES 149 is used to deduce the function of GR polynucleotides and GAM oligonucleotides which are "high" in the hierarchy, i.e. GAM oligonucleotides which modulate expression of other GR polynucleotides rather than directly modulating expression of target proteins. A preferred approach is as follows: The function of protein-modulating GAM oligonucleotides is deducible from the proteins which they modulate, provided that the function of these target proteins is known. The function of "higher" GAM oligonucleotides may be deduced by comparing the function of protein-modulating GAM oligonucleotides with the hierarchical relationships by which the "higher" GAM oligonu-

cleotides are connected to the protein-modulating GAM oligonucleotides. For example, given a group of several protein-modulating GAM oligonucleotides which collectively cause a protein expression pattern typical of a certain cell-type, then a "higher" GAM oligonucleotide is sought which modulates expression of GR polynucleotides which perhaps modulate expression of other GR polynucleotides which eventually modulate expression of the given group of protein-modulating GAM oligonucleotide. The "higher" GAM oligonucleotide found in this manner is taken to be responsible for differentiation of that cell-type, as per the conceptual model of the invention described hereinabove with reference to Fig. 6.

[0419] Reference is now made to Fig. 19 which is a block diagram illustrating different utilities of oligonucleotide of the novel group of oligonucleotides of the present invention referred to here as GAM oligonucleotides and GR polynucleotides. The present invention discloses a first plurality of novel oligonucleotides referred to here as GAM oligonucleotides and a second plurality of operon-like polynucleotides referred to here as GR polynucleotides, each of the GR polynucleotide encoding a plurality of GAM oligonucleotides. The present invention further discloses a

very large number of known target genes, which are bound by, and the expression of which is modulated by each of the novel oligonucleotides of the present invention. Published scientific data referenced by the present invention provides specific, substantial, and credible evidence that the above mentioned target genes modulated by novel oligonucleotides of the present invention, are associated with various diseases. Specific novel oligonucleotides of the present invention, target genes thereof and diseases associated therewith, are described hereinbelow with reference to Tables 1 through 13. It is therefore appreciated that a function of GAM oligonucleotides and GR polynucleotides of the present invention is modulation of expression of target genes related to known diseases, and that therefore utilities of novel oligonucleotides of the present invention include diagnosis and treatment of the above mentioned diseases.

[0420] Fig. 19 describes various types of diagnostic and therapeutic utilities of novel oligonucleotides of the present invention. A utility of novel oligonucleotide of the present invention is detection of GAM oligonucleotides and of GR polynucleotides. It is appreciated that since GAM oligonucleotides and GR polynucleotides modulate expression of



disease related target genes, that detection of expression of GAM oligonucleotides in clinical scenarios associated with said diseases is a specific, substantial and credible utility. Diagnosis of novel oligonucleotides of the present invention may preferably be implemented by RNA expression detection techniques, including but not limited to biochips, as is well known in the art. Diagnosis of expression of oligonucleotides of the present invention may be useful for research purposes, in order to further understand the connection between the novel oligonucleotides of the present invention and the above mentioned related diseases, for disease diagnosis and prevention purposes, and for monitoring disease progress.

[0421] Another utility of novel oligonucleotides of the present invention is anti-GAM therapy, a mode of therapy which allows up regulation of a disease-related target gene of a novel GAM oligonucleotide of the present invention, by lowering levels of the novel GAM oligonucleotide which naturally inhibits expression of that target gene. This mode of therapy is particularly useful with respect to target genes which have been shown to be under-expressed in association with a specific disease. Anti-GAM therapy is further discussed hereinbelow with reference to Figs. 20A

and 20B.

[0422] A further utility of novel oligonucleotides of the present invention is GAM replacement therapy, a mode of therapy which achieves down regulation of a disease related target gene of a novel GAM oligonucleotide of the present invention, by raising levels of the GAM which naturally inhibits expression of that target gene. This mode of therapy is particularly useful with respect to target genes which have been shown to be over-expressed in association with a specific disease. GAM replacement therapy involves introduction of supplementary GAM products into a cell, or stimulation of a cell to produce excess GAM products. GAM replacement therapy may preferably be achieved by transfecting cells with an artificial DNA molecule encoding a GAM which causes the cells to produce the GAM product, as is well known in the art.

[0423] Yet a further utility of novel oligonucleotides of the present invention is modified GAM therapy. Disease conditions are likely to exist, in which a mutation in a binding site of a GAM RNA prevents natural GAM RNA to effectively bind inhibit a disease related target gene, causing up regulation of that target gene, and thereby contributing to the disease pathology. In such conditions, a modi-

fied GAM oligonucleotides is designed which effectively binds the mutated GAM binding site, i.e. is an effective anti-sense of the mutated GAM binding site, and is introduced in disease effected cells. Modified GAM therapy is preferably achieved by transfecting cells with an artificial DNA molecule encoding the modified GAM which causes the cells to produce the modified GAM product, as is well known in the art.

[0424] An additional utility of novel GAM of the present invention is induced cellular differentiation therapy. An aspect of the present invention is finding oligonucleotides which determine cellular differentiation, as described hereinabove with reference to Fig. 18. Induced cellular differentiation therapy comprises transfection of cell with such GAM oligonucleotides thereby determining their differentiation as desired. It is appreciated that this approach may be widely applicable, inter alia as a means for auto transplantation harvesting cells of one cell-type from a patient, modifying their differentiation as desired, and then transplanting them back into the patient. It is further appreciated that this approach may also be utilized to modify cell differentiation in-vivo, by transfecting cells in a genetically diseased tissue with a cell-differentiation determin-

ing GAM thus stimulating these cells to differentiate appropriately.

[0425] Reference is now made to Figs. 20A and 20B, simplified diagrams which when taken together illustrate anti-GAM therapy mentioned hereinabove with reference to Fig. 19. A utility of novel GAMs of the present invention is anti-GAM therapy, a mode of therapy which allows up regulation of a disease-related target gene of a novel GAM of the present invention, by lowering levels of the novel GAM which naturally inhibits expression of that target gene. Fig. 20A shows a normal GAM inhibiting translation of a target gene by binding of GAM RNA to a BINDING SITE found in an untranslated region of GAM TARGET RNA, as described hereinabove with reference to Fig. 8.

[0426] Fig. 20B shows an example of anti-GAM therapy. ANTI-GAM RNA is short artificial RNA molecule the sequence of which is an anti-sense of GAM RNA. Anti-GAM treatment comprises transfecting diseased cells with ANTI-GAM RNA, or with a DNA encoding thereof. The ANTI-GAM RNA binds the natural GAM RNA, thereby preventing binding of natural GAM RNA to its BINDING SITE. This prevents natural translation inhibition of GAM TARGET RNA by GAM RNA, thereby up regulating expression of GAM TARGET

## PROTEIN.

[0427] It is appreciated that anti-GAM therapy is particularly useful with respect to target genes which have been shown to be under-expressed in association with a specific disease.

[0428] Furthermore, anti-GAM therapy is particularly useful, since it may be used in situations in which technologies known in the art as RNAi and siRNA can not be utilized. As in known in the art, RNAi and siRNA are technologies which offer means for artificially inhibiting expression of a target protein, by artificially designed short RNA segments which bind complementarily to mRNA of said target protein. However, RNAi and siRNA can not be used to directly up regulate translation of target proteins.

[0429] Reference is now made to Fig. 21A, which is a bar graph illustrating performance results of the hairpin detector 114 (Fig. 9) constructed and operative in accordance with a preferred embodiment of the present invention.

[0430] Fig. 21A illustrates efficacy of several features used by the hairpin detector 114 to detect GAM FOLDED PRECURSOR RNAs (Fig. 8). The values of each of these features is compared between a set of published miRNA precursor oligonucleotides, represented by shaded bars, and a set of random hairpins folded from the human genome denoted

hereinbelow as a hairpin background set, represented by white bars. The published miRNA precursor oligonucleotides set is taken from RFAM database, Release 2.1 and includes 148 miRNA oligonucleotides from H.Sapiens. The background set comprises a set of 10,000 hairpins folded from the human genome.

[0431] It is appreciated that the hairpin background set is expected to comprise some valid, previously undetected hairpin-shaped miRNA precursor-like GAM FOLDED PRE-CURSOR RNAs of the present invention, and many hairpin-shaped sequences that are not hairpin-shaped miRNA-like precursors.

[0432] For each feature, the bars depict the percent of known miRNA hairpin precursors (shaded bars) and the percent of background hairpins (white bars) that pass the threshold for that feature. The percent of known miRNA oligonucleotides that pass the threshold indicates the sensitivity of the feature, while the corresponding background percent implies the specificity of the feature, although not precisely, because the background set comprises both true and false examples.

[0433] The first bar pair, labeled Thermodynamic Stability Selection, depicts hairpins that have passed the selection of

"families" of closely related hairpin structures, as described hereinabove with reference to Fig. 12B.

[0434] The second bar pair, labeled Hairpin Score, depicts hairpins that have been selected by hairpin detector 114 (Fig. 12B), regardless of the families selection.

[0435] The third bar pair, labeled Conserved, depicts hairpins that are conserved in human, mouse and rat, (UCSC Goldenpath (TM) HG16 database).

[0436] The fourth bar pair, labeled Expressed, depicts hairpins that are found in EST blocks.

[0437] The fifth bar pair, labeled Integrated Selection, depicts hairpin structures predicted by a preferred embodiment of the present invention to be valid GAM PRECURSORS. In a preferred embodiment of the present invention, a hairpin may be considered to be a GAM PRECURSOR if its hairpin detector score is above 0, and it is in one of the following groups: a) in an intron and conserved or b) in an intergenic region and conserved or c) in an intergenic region and expressed, as described below. Further filtering of GAM precursor may be obtained by selecting hairpins with a high score of Dicer-cut location detector 116 as described hereinabove with reference to Figs. 13A-13C, and with predicted miRNA oligonucleotides, which pass the

low complexity filter as described hereinabove, and whose targets are selected by the target gene binding site detector 118 as described hereinabove with reference to Figs. 14A–14B.

[0438] It is appreciated that these results validate the sensitivity and specificity of the hairpin detector 114 (Fig. 9) in identifying novel GAM FOLDED PRECURSOR RNAs, and in effectively distinguishing them from the abundant hairpins found in the genome.

[0439] Reference is now made to Fig. 21B, which is a line graph illustrating accuracy of a Dicer-cut location detector 116 (Fig. 9) constructed and operative in accordance with a preferred embodiment of the present invention.

[0440] To determine the accuracy of the Dicer-cut location detector 116, a stringent training and test set was chosen from the abovementioned set of 440 known miRNA oligonucleotides, such that no two miRNA oligonucleotides in the set are homologous. This was performed to get a lower bound on the accuracy and avoid effects of similar known miRNA oligonucleotides appearing in both the training and test sets. On this stringent set of size 204, mfold cross validation with  $k=3$  was performed to determine the percent of known miRNA oligonucleotides



in which the dicer-cut location detector 116 described hereinabove predicted the correct miRNA oligonucleotide up to two nucleotides from the correct location. The accuracy of the TWO PHASED predictor is depicted in the graph. The accuracy of the first phase of the TWO PHASED predictor is depicted by the upper line, and that of both phases of the TWO PHASED predictor is depicted by the lower line. Both are binned by the predictor score, where the score is the score of the first stage.

[0441] It is appreciated that these results validate the accuracy of the Dicer-cut location detector 116.

[0442] Reference is now made to Fig. 21C, which is a bar graph illustrating the performance results of the target gene binding site detector 118 (Fig. 14A) constructed and operative in accordance with a preferred embodiment of the present invention.

[0443] Fig. 21C illustrates specificity and sensitivity of the target gene binding site detector 118. The values presented are the result of testing 10000 artificial miRNA oligonucleotide sequences (random 22 nt sequences with the same base composition as published miRNA oligonucleotide sequence). Adjusting the threshold parameters to fulfill 90% sensitivity of validated, published miRNA-3'UTR

pairs, requires the P VAL of potential target gene sequences–Dicer–cut sequences to be less than 0.01 and also the P VAL of potential target ortholog gene sequences–Dicer–cut sequences to be less than 0.05. The target gene binding site detector 118 can filter out 99.7% of potential miRNA/gene pairs, leaving only the 0.3% that contain the most promising potential miRNA/gene pairs. Limiting the condition for the P VAL of potential target ortholog gene sequences–Dicer–cut sequences to be less than 0.01 reduces the sensitivity ratio to 70% but filters out more than 50% of the remaining 0.3%, to a final ratio of less than 0.15%.

[0444] It is appreciated that these results validate the sensitivity and specificity of the target gene binding site detector 118.

[0445] Reference is now made to Fig. 22, which is a summary table of laboratory results validating the expression of 29 novel human GAM RNA oligonucleotides in HeLa cells or, alternatively, in liver or thymus tissues detected by the bioinformatic oligonucleotide detection engine 100 (Fig. 9).

[0446] As a positive control, we used a reference set of eight known human miRNA oligonucleotides: hsa–MIR–21; hsa–

MIR-27b; hsa-MIR-186; hsa-MIR-93; hsa-MIR-26a; hsa-MIR-191; hsa-MIR-31; and hsa-MIR-92. All positive controls were successfully validated by sequencing.

[0447] The table of Fig. 22 lists all GAM RNA predictions whose expression was validated. The field "Primer Sequence" contains the "specific" part of the primer; the field "Sequenced sequence" represents the nucleotide sequence detected by cloning (excluding the hemispecific primer sequence); the field "Predicted GAM RNA" contains the GAM RNA predicted sequence; the field "Distance indicate the distance from Primer; the number of mismatches between the "specific" region of the primer and the corresponding part of the GAM RNA sequence; the field "GAM Name" contains GAM RNA PRECURSOR ID followed by "A" or "B", which represents the GAM RNA position on the precursor as elaborated in the attached Tables.

[0448] A primer was designed such that its first half, the 5' region, is complementary to the adaptor sequence and its second half, the 3' region, anneals to the 5' terminus of GAM RNA sequence, yielding a hemispecific primer (as elaborated hereinbelow in the Methods section). A sample of 13 predicted GAM RNA sequences was examined by PCR using hemispecific primers and a primer specific to

the 3' adaptor. PCR products were cloned into plasmid vectors and then sequenced. For all 13 predicted GAM RNA sequences, the GAM RNA sequence found in the hemispecific primer plus the sequence observed between the hemispecific primer and the 3' adaptor was completely included in the expected GAM RNA sequence (rows 1–7, and 29). The rest are GAM RNA predictions that were verified by cloning and sequencing, yet, by using a primer that was originally designed for a slightly different prediction.

[0449] It is appreciated that failure to detect a predicted oligonucleotide in the lab does not necessarily indicate a mistaken bioinformatic prediction. Rather, it may be due to technical sensitivity limitation of the lab test, or because the predicted oligonucleotides are not expressed in the tissue examined, or at the development phase tested. The observed GAM RNAs may be strongly expressed in HeLa cells while the original GAM RNAs are expressed at low levels in HeLa cells or not expressed at all. Under such circumstances, primer sequences containing up to three mismatches from a specific GAM RNA sequence may amplify it. Thus, we also considered cases in which differences of up to 3 mismatches in the hemispecific primer

occur.

[0450] The 3' terminus of observed GAM RNA sequences is often truncated or extended by one or two nucleotides. Cloned sequences that were sequenced from both 5' and 3' termini have an asterick appended to the row number.

[0451] Interestingly, the primer sequence followed by the observed cloned sequence is contained within five GAM RNA sequences of different lengths, and belong to 24 precursors derived from distinct loci (Row 29). Out of these, one precursor appears four times in the genome and its corresponding GAM Names are 351973-A, 352169-A, 352445-A and 358164-A.

[0452] The sequence presented in Row 29 is a representative of the group of five GAM RNAs. The full list of GAM RNA sequences and their corresponding precursors is as follows (each GAM RNA sequence is followed by the GAM Name):  
TCACTGCAACCTCCACCTCCCA (352092,  
352651,355761),TCACTGCAACCTCCACCTCCCG (351868,  
352440, 351973, 352169, 352445, 358164, 353737,  
352382, 352235, 352232, 352268, 351919, 352473,  
352444, 353638, 353004, 352925, 352943), TCACTG-  
CAACCTCCACCTCCTG  
(358311),TCACTGCAACCTCCACCTTCAG (353323), and

TCACTGCAACCTCCACCTTCCG (353856).

[0453] **METHOD SECTION**

[0454] **CELL LINES**

[0455] Three common human cell lines, obtained from Dr. Yonat Shemer at Soroka Medical Center, Be'er Sheva, Israel, were used for RNA extraction; Human Embryonic Kidney HEK-293 cells, Human Cervix Adenocarcinoma HeLa cells and Human Prostate Carcinoma PC3 cells.

[0456] **RNA PURIFICATION**

[0457] Several sources of RNA were used to prepare libraries:

[0458] Total HeLa S100 RNA was prepared from HeLa S100 cellular fraction (4C Biotech, Belgium) through an SDS (1%)–Proteinase K (200g/ml) 30 minute incubation at 37C followed by an acid Phenol–Chloroform purification and isopropanol precipitation (Sambrook et al; Molecular Cloning– A Laboratory Manual).

[0459] Total HeLa, HEK-293 and PC3 cell RNA was prepared using the standard Tri–Reagent protocol (Sigma) according to the manufacturer's instructions, except that 1 volume of isopropanol was substituted with 3 volumes of ethanol.

[0460] Nuclear and Cytoplasmic RNA was prepared from HeLa or HEK-293 cells in the following manner:

[0461] Cells were washed and harvested in ice-cold PBS and precipitated in a swing-out rotor at 1200 rpm at 4°C for 5 minutes. Pellets were loosened by gentle vortexing. 4 ml of "NP40 lysis buffer" (10 mM TrisHCl, 5 mM MgCl<sub>2</sub>, 10 mM NaCl, 0.5% Nonidet P40, 1 mM Spermidine, 1 mM DTT, 140 U/ml rNasase) was then added per 5 × 10<sup>7</sup> cells. Cells and lysis buffer were incubated for 5 minutes on ice and centrifuged in a swing-out rotor at 500 × g at 4°C for 5 minutes. Supernatant, termed cytoplasm, is carefully removed to a tube containing SDS (1% final) and proteinase-K (200 g/ml final). Pellet, termed nuclear fraction, is re-washed and incubated with a similar amount of fresh lysis buffer. Lysis is monitored visually under a microscope at this stage, typically for 5 minutes. Nuclei are pelleted in a swing-out rotor at 500 × g at 4°C for 5 minutes. Supernatant is pooled, incubated at 37°C for 30 minutes, Phenol/Chloroform-extracted, and RNA is alcohol-precipitated (Sambrook et al). Nuclei are loosened and then homogenized immediately in >10 volumes of Tri-Reagent (Sigma). Nuclear RNA is then prepared according to the manufacturer's instructions.

[0462] TOTAL TISSUE RNA

[0463] Total tissue RNA was obtained from Ambion USA, and in-

cluded Human Liver, Thymus, Placenta, Testes and Brain.

#### [0464] RNA SIZE FRACTIONATION

[0465] RNA used for libraries was always size-fractionated. Fractionation was done by loading up to 500 g RNA per YM100 Amicon Microcon column (Millipore) followed by a 500xg centrifugation for 40 minutes at 4C. Flow-through "YM100" RNA is about one quarter of the total RNA and was used for library preparation or fractionated further by loading onto a YM30 Amicon Microcon column (Millipore) followed by a 13,500xg centrifugation for 25 minutes at 4C. Flow-through "YM30" was used for library preparation "as is" and consists of less than 0.5% of total RNA. Additional size fractionation was achieved during library preparation.

#### [0466] LIBRARY PREPARATION

[0467] Two types of cDNA libraries, designated "One-tailed" and "Ligation", were prepared from the one of the abovementioned fractionated RNA samples. RNA was dephosphorylated and ligated to an RNA (designated with lowercase letters)-DNA (designated with UPPERCASE letters) hybrid 5'-phosphorylated, 3' idT blocked 3'-adapter (5'-P-uuuAACCGCATCCTTCTC-idT-3' Dharmacon # P-



002045-01-05) (as elaborated in Elbashir et al., *Genes Dev.* 15:188-200 (2001)) resulting in ligation only of RNase III type cleavage products. 3'-Ligated RNA was excised and purified from a half 6%, half 13% polyacrylamide gel to remove excess adapter with a Nanosep 0.2M centrifugal device (Pall) according to instructions, and precipitated with glycogen and 3 volumes of ethanol. Pellet was resuspended in a minimal volume of water.

- [0468] For the "Ligation" library, a DNA (UPPERCASE)-RNA (lowercase) hybrid 5'-adapter (5'-TACTAATACGACTCACTaaa-3' Dharmacon # P-002046-01-05) was ligated to the 3'-adapted RNA, reverse transcribed with "EcoRI-RT": (5'-GACTAGCTGGAATTCAAGGATGCGGTAAA-3'), PCR amplified with two external primers essentially as in Elbashir et al. (2001), except that primers were "EcoRI-RT" and "PstI Fwd"(5'-CAGCCAACGCTGCAGATACGACTCACTAAA-3'). This PCR product was used as a template for a second round of PCR with one hemispecific and one external primer or with two hemispecific primers.
- [0469] For the "One-tailed" library, the 3'-adapted RNA was annealed to 20pmol primer "EcoRI RT" by heating to 70C and

cooling 0.1C/sec to 30C and then reverse-transcribed with Superscript II RT (according to manufacturer's instructions, Invitrogen) in a 20l volume for 10 alternating 5 minute cycles of 37C and 45C. Subsequently, RNA was digested with 1l 2M NaOH and 2mM EDTA at 65C for 10 minutes. cDNA was loaded on a polyacrylamide gel, excised and gel-purified from excess primer as above (invisible, judged by primer run alongside) and resuspended in 13l of water. Purified cDNA was then oligo-dC tailed with 400U of recombinant terminal transferase (Roche Molecular Biochemicals), 1l 100M dCTP, 1l 15mM CoCl<sub>2</sub>, and 4l reaction buffer, to a final volume of 20l for 15 minutes at 37C. Reaction was stopped with 2l 0.2M EDTA and 15l 3M NaOAc pH 5.2. Volume was adjusted to 150l with water, Phenol: Bromochloropropane 10:1 extracted and subsequently precipitated with glycogen and 3 volumes of ethanol. C-tailed cDNA was used as a template for PCR with the external primers "T3-PstBsg(G/I)18"(5'-AATTAACCCTCACTAAAGGCTGCAG GTGCAGGIGGGIIGGGIIGGGIIGN-3' where I stands for Inosine and N for any of the 4 possible deoxynucleotides), and with "EcoRI Nested"(5'-GGAATTCAAGGATGCGGTTA-3'). This PCR

product was used as a template for a second round of PCR with one hemispecific and one external primer or with two hemispecific primers.

#### [0470] PRIMER DESIGN AND PCR

[0471] Hemispecific primers were constructed for each predicted GAM RNA oligonucleotide by an in-house program designed to choose about half of the 5' or 3' sequence of the GAM RNA corresponding to a TM of about 30–34C constrained by an optimized 3' clamp, appended to the cloning adapter sequence (for "One-tailed" libraries, 5'-GGNNGGGNNG on the 5' end or TTTAACCGCATC-3' on the 3' end of the GAM RNA; for "Ligation" libraries, the same 3' adapter and 5'-CGACTCACTAAA on the 5' end of the GAM RNA). Consequently, a fully complementary primer of a TM higher than 60C was created covering only one half of the GAM RNA sequence permitting the unbiased elucidation by sequencing of the other half.

[0472] For each primer, the following criteria were used: Primers were graded according to the TM of the primer half and the nucleotide content of 3 nucleotides of the 3' clamp from worst to best, roughly: GGG-3' < CCC-3' < TTT-3'/AAA-3' < GG-3' < CC-3' < a TM lower than 30 < a TM higher than 34 < TT-3'/AA-3' < 3G/C nucleotide

combination <3 A/T nucleotide combination <any combination of two/three different nucleotides <any combination of three/three different nucleotides.

[0473] VALIDATION PCR PRODUCT BY SOUTHERN BLOT

[0474] GAM RNA oligonucleotides were validated by hybridization of Polymerase Chain Reaction (PCR)-product Southern blots with a probe to the predicted GAM RNA.

[0475] PCR product sequences were confirmed by Southern blot (Southern E.M., Biotechnology 1992,24:122-139 (1975)) and hybridization with DNA oligonucleotide probes synthesized as complementary (antisense) to predicted GAM RNA oligonucleotides. Gels were transferred onto a Bio-dyne PLUS 0.45m (Pall) positively charged nylon membrane and UV cross-linked. Hybridization was performed overnight with DIG-labeled probes at 42°C in DIG Easy-Hyb buffer (Roche). Membranes were washed twice with 2xSSC and 0.1% SDS for 10 minutes at 42°C and then washed twice with 0.5xSSC and 0.1% SDS for 5 min at 42°C. The membrane was then developed by using a DIG luminescent detection kit (Roche) using anti-DIG and CSPD reaction, according to the manufacturer's protocol. All probes were prepared according to the manufacturer's (Roche Molecular Biochemicals) protocols: Digoxigenin

(DIG) labeled antisense transcripts were prepared from purified PCR products using a DIG RNA labeling kit with T3 RNA polymerase. DIG-labeled PCR was prepared by using a DIG PCR labeling kit. 3'-DIG-tailed oligo ssDNA antisense probes, containing DIG-dUTP and dATP at an average tail length of 50 nts were prepared from 100pmole oligonucleotides with the DIG Oligonucleotide Labeling Kit. Control reactions contained all of the components of the test reaction except library template.

#### [0476] VALIDATION OF PCR PRODUCT BY NESTED PCR ON THE LIGATION

[0477] To further validate predicted GAM PCR product sequence derived from hemi-primers, a PCR-based diagnostic technique was devised to amplify only those products containing at least two additional nucleotides of the non hemi-primer defined part of the predicted GAM RNA oligonucleotide. In essence, a diagnostic primer was designed so that its 3' end, which is the specificity determining side, was identical to the desired GAM RNA oligonucleotide, 2-10 nts (typically 4-7, chosen for maximum specificity) further into its 3' end than the nucleotide stretch primed by the hemi-primer. The hemi-primer PCR product was first ligated into a T-cloning vector (pTZ57/T or pGEM-T)

as described hereinabove. The ligation reaction mixture was used as template for the diagnostic PCR under strict annealing conditions with the new diagnostic primer in conjunction with a general plasmid-homologous primer, resulting in a distinct ~200 base-pair product. This PCR product can be directly sequenced, permitting the elucidation of the remaining nucleotides up to the 3' of the mature GAM RNA oligonucleotide adjacent to the 3' adapter. Alternatively, following analysis of the diagnostic PCR reaction on an agarose gel, positive ligation reactions (containing a band of the expected size) were transformed into *E. coli*. Using this same diagnostic technique and as an alternative to screening by Southern blot colony hybridization, transformed bacterial colonies were screened by colony-PCR (Gussow, D. and Clackson, T, *Nucleic Acids Res.* 17:4000 (1989)) with the nested primer and the vector primer, prior to plasmid purification and sequencing.

[0478] VALIDATION OF PCR PRODUCT BY CLONING AND SEQUENCING

[0479] CLONE SEQUENCING: PCR products were inserted into pGEM-T (Promega) or pTZ57/T (MBI Fermentas), heat-shock transformed into competent JM109 *E. coli* (Promega) and seeded on LB-Ampicilin plates with IPTG

and Xgal. White and light blue colonies were transferred to duplicate gridded plates, one of which was blotted onto a membrane (Biodyme Plus, Pall) for hybridization with DIG tailed oligo probes (according to instructions, Roche) complementary to the expected GAM. Plasmid DNA from positive colonies was sequenced.

[0480] It is appreciated that the results summarize in Fig. 22 validate the efficacy of the bioinformatic oligonucleotide detection engine 100 of the present invention.

[0481] Reference is now made to Fig. 23A, which is a schematic representation of a novel human GR polynucleotide, located on chromosome 9, comprising 2 known human MIR oligonucleotides – MIR24 and MIR23, and 2 novel GAM oligonucleotides, herein designated GAM7617 and GAM252 (later discovered by other researchers as hsa-mir-27b), all marked by solid black boxes. Fig. 23A also schematically illustrates 6 non-GAM hairpin sequences, and one non-hairpin sequence, all marked by white boxes, and serving as negative controls. By "non-GAM hairpin sequences" is meant sequences of a similar length to known MIR PRECURSOR sequences, which form hairpin secondary folding pattern similar to MIR PRECURSOR hairpins, and yet which are assessed by the bioinformatic

oligonucleotide detection engine 100 not to be valid GAM PRECURSOR hairpins. It is appreciated that Fig. 23A is a simplified schematic representation, reflecting only the order in which the segments of interest appear relative to one another, and not a proportional distance between the segments.

[0482] Reference is now made to Fig. 23B, which is a schematic representation of secondary folding of each of the MIRs and GAMs of the GR MIR24, MIR23, GAM7617 and GAM252, and of the negative control non-GAM hairpins, herein designated N2, N3, N252, N4, N6 and N7. N0 is a non-hairpin control, of a similar length to that of known MIR PRECURSOR hairpins. It is appreciated that the negative controls are situated adjacent to and in between real MIR oligonucleotides and GAM predicted oligonucleotides and demonstrates similar secondary folding patterns to that of known MIRs and GAMs.

[0483] Reference is now made to Fig. 23C, which is a picture of laboratory results of a PCR test upon a YM100 size-fractionated "ligation"-library, utilizing a set of specific primer pairs located directly inside the boundaries of the hairpins. Due to the nature of the library the only PCR amplifiable products can result from RNaseIII type enzyme



cleaved RNA, as expected for legitimate hairpin precursors presumed to be produced by DROSHA (Lee et al, Nature 425 415–419, 2003). Fig. 23C demonstrates expression of hairpin precursors of known MIR oligonucleotides – hsa-mir23 and hsa-mir24, and of novel bioinformatically-detected GAM7617 and GAM252 hairpins predicted bioinformatically by a system constructed and operative in accordance with a preferred embodiment of the present invention. Fig. 23C also shows that none of the 7 controls (6 hairpins designated N2, N3, N23, N4, N6 and N7 and 1 non-hairpin sequence designated N0) were expressed. N252 is a negative control sequence partially overlapping GAM252.

[0484] In the picture, test lanes including template are designated "+" and the control lane is designated "-". The control reaction contained all the components of the test reaction except library template. It is appreciated that for each of the tested hairpins, a clear PCR band appears in the test ("+") lane, but not in the control ("-") lane.

[0485] Figs. 23A through 23C, when taken together validate the efficacy of the bioinformatic oligonucleotide detection engine in: (a) detecting known MIR oligonucleotides; (b) detecting novel GAM PRECURSOR hairpins which are found

adjacent to these MIR oligonucleotides, and which despite exhaustive prior biological efforts and bioinformatic detection efforts, went undetected; (c) discerning between GAM (or MIR) PRECURSOR hairpins, and non-GAM hairpins.

[0486] It is appreciated that the ability to discern GAM-hairpins from non-GAM-hairpins is very significant in detecting GAM oligonucleotides since hairpins are highly abundant in the genome. Other MIR prediction programs have not been able to address this challenge successfully.

[0487] Reference is now made to Fig. 24A which is an annotated sequence of an EST comprising a novel GAM oligonucleotides detected by the oligonucleotide detection system of the present invention. Fig. 24A shows the nucleotide sequence of a known human non-protein-coding EST (Expressed Sequence Tag), identified as EST72223. The EST72223 clone obtained from TIGR database (Kirkness and Kerlavage, 1997) was sequenced to yield the above 705bp transcript with a polyadenyl tail. It is appreciated that the sequence of this EST comprises sequences of one known miRNA oligonucleotide, identified as hsa-MIR98, and of one novel GAM oligonucleotide referred to here as GAM25, detected by the bioinformatic oligonu-

cleotide detection engine 100 (Fig. 9) of the present invention.

[0488] The sequences of the precursors of the known MIR98 and of the predicted GAM25 precursors are marked in bold, the sequences of the established miRNA 98 and of the predicted miRNA-like oligonucleotide GAM25 are underlined.

[0489] Reference is now made to Figs. 24B, 24C and 24D that are pictures of laboratory results, which when taken together demonstrate laboratory confirmation of expression of the bioinformatically-detected novel oligonucleotide of Fig. 24A. In two parallel experiments, an enzymatically synthesized capped, EST72223 RNA transcript, was incubated with Hela S100 lysate for 0 minutes, 4 hours and 24 hours. RNA was subsequently harvested, run on a denaturing polyacrylamide gel, and reacted with either a 102 nt antisense MIR98 probe or a 145 nt antisense GAM25 precursor transcript probe respectively. The Northern blot results of these experiments demonstrated processing of EST72223 RNA by Hela lysate (lanes 2–4, in Figs. 24B and 24C), into ~80bp and ~22bp segments, which reacted with the MIR98 precursor probe (Fig. 24B), and into ~100bp and ~24bp segments, which reacted with the

GAM25 precursor probe (Fig. 24C). These results demonstrate the processing of EST72223 by Hela lysate into MIR98 precursor and GAM25 precursor. It is also appreciated from Fig. 24C (lane 1) that Hela lysate itself reacted with the GAM25 precursor probe, in a number of bands, including a ~100bp band, indicating that GAM25-precursor is endogenously expressed in Hela cells. The presence of additional bands, higher than 100bp in lanes 5-9 probably corresponds to the presence of nucleotide sequences in Hela lysate, which contain the GAM25 sequence.

[0490] In addition, in order to demonstrate the kinetics and specificity of the processing of MIR98 and GAM25 precursors into their respective mature, "diced" segments, transcripts of MIR98 and of the bioinformatically predicted GAM25 precursors were similarly incubated with Hela S100 lysate, for 0 minutes, 30 minutes, 1 hour and 24 hours, and for 24 hours with the addition of EDTA, added to inhibit Dicer activity, following which RNA was harvested, run on a polyacrylamide gel and reacted with MIR98 and GAM25 precursor probes. Capped transcripts were prepared for in-vitro RNA cleavage assays with T7 RNA polymerase, including a m7G(5')ppp(5')G-capping re-

action using the T7-mMessage mMachine kit (Ambion). Purified PCR products were used as template for the reaction. These were amplified for each assay with specific primers containing a T7 promoter at the 5' end and a T3 RNA polymerase promoter at the 3' end. Capped RNA transcripts were incubated at 30C in supplemented, dialysis concentrated, Hela S100 cytoplasmic extract (4C Biotech, Seneffe, Belgium). The Hela S100 was supplemented by dialysis to a final concentration of 20mM Hepes, 100mM KCl, 2.5mM MgCl<sub>2</sub>, 0.5mM DTT, 20% glycerol and protease inhibitor cocktail tablets (Complete mini Roche Molecular Biochemicals). After addition of all components, final concentrations were 100mM capped target RNA, 2mM ATP, 0.2mM GTP, 500U/ml RNasin, 25g/ml creatine kinase, 25mM creatine phosphate, 2.5mM DTT and 50% S100 extract. Proteinase K, used to enhance Dicer activity (Zhang et al., EMBO J. 21, 5875–5885 (2002)) was dissolved in 50mM Tris-HCl pH 8, 5mM CaCl<sub>2</sub>, and 50% glycerol, was added to a final concentration of 0.6 mg/ml. Cleavage reactions were stopped by the addition of 8 volumes of proteinase K buffer (200mM Tris-HCl, pH 7.5, 25mM EDTA, 300mM NaCl, and 2% SDS) and incubated at 65C for 15min at different time points

(0, 0.5, 1, 4, 24h) and subjected to phenol/chloroform extraction. Pellets were dissolved in water and kept frozen. Samples were analyzed on a segmented half 6%, half 13% polyacrylamide 1XTBE-7M Urea gel.

[0491] The Northern blot results of these experiments demonstrated an accumulation of a ~22bp segment which reacted with the MIR98 precursor probe, and of a ~24bp segment which reacted with the GAM25 precursor probe, over time (lanes 5–8). Absence of these segments when incubated with EDTA (lane 9), which is known to inhibit Dicer enzyme (Zhang et al., 2002), supports the notion that the processing of MIR98 and GAM25 precursors into their "diced" segments is mediated by Dicer enzyme, found in Hela lysate. Other RNases do not utilize divalent cations and are thus not inhibited by EDTA. The molecular sizes of EST72223, MIR-98 and GAM25 and their corresponding precursors are indicated by arrows.

[0492] Fig. 24D present Northern blot results of same above experiments with GAM25 probe (24 nt). The results clearly demonstrated the accumulation of mature GAM25 oligonucleotide after 24 h.

[0493] To validate the identity of the band shown by the lower arrow in figs. 24C and 24D, a RNA band parallel to a

marker of 24 base was excised from the gel and cloned as in Elbashir et al (2001) and sequenced. 90 clones corresponded to the sequence of mature GAM25 oligonucleotide, three corresponded to GAM25\* (the opposite arm of the hairpin with a 1–3 nt 3' overhang) and two to the hairpin–loop.

[0494] GAM25 was also validated endogenously by sequencing from both sides from a HeLa YM100 total-RNA "ligation" libraries, utilizing hemispecific primers as described in Fig. 22.

[0495] Taken together, these results validate the presence and processing of a novel MIR-like oligonucleotide, GAM25, which was predicted bioinformatically. The processing of this novel GAM oligonucleotide product, by HeLa lysate from EST72223, through its precursor, to its final form was similar to that observed for known miRNA oligonucleotide, MIR98.

[0496] Transcript products were 705 nt (EST72223), 102 nt (MIR98 precursor), 125 nt (GAM25 precursor) long. EST72223 was PCR amplified with T7-EST 72223 forward primer:

5'-TAATACGACTCACTATAGGCCCTTATTAGAGGATTCTGCT  
-3' and T3-EST72223 reverse

primer:"-AATTAACCCTCACTAAAGGTTTTTTTTTCCTGAGA  
CAGAGT-3'.MIR98 was PCR amplified using EST72223 as  
a template with T7MIR98 forward primer:

5'-TAATACGACTCACTATAGGGTGAGGTAGTAAGTTGTATT  
GTT-3'and T3MIR98 reverse primer:

5'-AATTAACCCTCACTAAAGGGAAAGTAGTAAGTTGTATAG  
TT-3'.GAM25 was PCR amplified using EST72223 as a  
template with GAM25 forward primer:

5'-GAGGCAGGAGAATTGCTTGA-3' and T3-EST72223 re-  
verse

primer:5'-AATTAACCCTCACTAAAGGCCTGAGACAGAGTCT  
TGCTC-3'.

[0497] It is appreciated that the data presented in Figs. 24A, 24B,  
24C and 24D when taken together validate the function of  
the bioinformatic oligonucleotide detection engine 100 of  
Fig. 9. Fig. 24A shows a novel GAM oligonucleotide bioin-  
formatically-detected by the bioinformatic oligonucleotide  
detection engine 100, and Figs. 24C and 24D show labo-  
ratory confirmation of the expression of this novel  
oligonucleotide. This is in accord with the engine training  
and validation methodology described hereinabove with  
reference to Fig. 9.

[0498] Reference is now made to Figs. 25A-C, which schemati-



cally represent three methods that are employed to identify GAM FOLDED PRECURSOR RNA from libraries. Each method involves the design of specific primers for PCR amplification followed by sequencing. The libraries include hairpins as double-stranded DNA with two different adaptors ligated to their 5' and 3' ends.

[0499] Reference is now made to Fig. 25A, which depicts a first method that uses primers designed to the stems of the hairpins. Since the stem of the hairpins often has bulges, mismatches, as well as G-T pairing, which is less significant in DNA than is G-U pairing in the original RNA hairpin, the primer pairs were engineered to have the lowest possible match to the other strand of the stem. Thus, the F-Stem primer, derived from the 5' stem region of the hairpin, was chosen to have minimal match to the 3' stem region of the same hairpin. Similarly, the R-stem primer, derived from the 3' region of the hairpin (reverse complementary to its sequence), was chosen to have minimal match to the 5' stem region of the same hairpin. The F-Stem primer was extended in its 5' sequence with the T3 primer (5'-ATTAACCCTCACTAAAGGGA-3') and the R-Stem primer was extended in its 5' sequence with the T7 primer (5'-TAATACGACTCACTATAGGG). The extension is

needed to obtain a large enough fragment for direct sequencing of the PCR product. Sequence data from the amplified hairpins is obtained in two ways. One way is the direct sequencing of the PCR products using the T3 primer that matches the extension of the F-Stem primer. Another way is the cloning of the PCR products into a plasmid, followed by PCR screening of individual bacterial colonies using a primer specific to the plasmid vector and either the R-Loop (Fig. 25B) or the F-Loop (Fig. 25C) primer. Positive PCR products are then sent for direct sequencing using the vector-specific primer.

[0500] Reference is now made to Fig. 25B, which depicts a second method in which R-Stem primer and R-Loop primers are used in a nested-PCR approach. First, PCR is performed with the R-Stem primer and the primer that matches the 5' adaptor sequence (5-ad primer). PCR products are then amplified in a second PCR using the R-Loop and 5-ad primers. As mentioned hereinabove, sequence data from the amplified hairpins is obtained in two ways. One way is the direct sequencing of the PCR products using the 5-ad primer. Another way is the cloning of the PCR products into a plasmid, followed by PCR screening of individual bacterial colonies using a primer specific

to the plasmid vector and F-Stem primer. Positive PCR products are then sent for direct sequencing using the vector-specific primer. It should be noted that optionally an extended R-Loop primer is designed that includes a T7 sequence extension, as described hereinabove (Fig. 25A) for the R-Stem primer. This is important in the first sequencing option in cases where the PCR product is too short for sequencing.

[0501] Reference is now made to Fig. 25C, which depicts a third method, which is the exact reverse of the second method described hereinabove (Fig. 25B). F-Stem and F-Loop primers are used in a nested-PCR approach. First, PCR is performed with the F-Stem primer and the primer that matches the 3' adaptor sequence (3-ad primer). PCR products are then amplified in a second PCR using the F-Loop and 3-ad primers. As in the other two methods, sequence data from the amplified hairpins is obtained in two ways. One way is the direct sequencing of the PCR products using the F-Loop primer. Another way is the cloning of the PCR products into a plasmid, followed by PCR screening of individual bacterial colonies using a primer specific to the plasmid vector and R-Stem primer. Positive PCR products are then sent for direct sequencing using

the vector-specific primer. It should be noted that optionally an extended F-Loop primer is designed that includes a T3 sequence extension, as described hereinabove (Fig. 25A) for the F-Stem primer. This is important in the first sequencing option in cases where the PCR product is too short for sequencing and also in order to enable the use of T3 primer.

[0502] In an embodiment of the present invention, the three methods mentioned hereinabove may be employed to validate the expression of GAM FOLDED PRECURSOR RNA.

[0503] Reference is now made to Fig. 26A, which is a flow chart with a general description of the design of the microarray to identify expression of published miRNA oligonucleotides, and of novel GAM oligonucleotides of the present invention.

[0504] A microarray that identifies miRNA oligonucleotides is designed (Fig. 26B). The DNA microarray is prepared by Agilent according to their SurePrint Procedure (reference describing their technology can be obtained from the Agilent website [www.agilent.com](http://www.agilent.com)). In this procedure, the oligonucleotide probes are synthesized on the glass surface. Other methods can also be used to prepare such microarray including the printing of pre-synthesized oligonu-

cleotides on glass surface or using the photolithography method developed by Affymetrix (Lockhart DJ et al., Nat Biotechnol.14:1675–1680 (1996)). The 60–mer sequences from the design are synthesized on the DNA microarray. The oligonucleotides on the microarray, termed "probes" are of the exact sequence as the designed 60–mer sequences. Importantly, the 60–mer sequences and the probes are in the sense orientation with regards to the miRNA oligonucleotides. Next, a cDNA library is created from size–fractionated RNA, amplified, and converted back to RNA (Fig. 26C). The resulting RNA is termed "cRNA". The conversion to RNA is done using a T7 RNA polymerase promoter found on the 3' adaptor (Fig. 26C; T7 NcoI–RNA–DNA 3'Adaptor). Since the conversion to cRNA is done in the reverse direction compared to the orientation of the miRNA oligonucleotides, the cRNA is reverse complementary to the probes and is able to hybridize to it. This amplified RNA is hybridized with the microarray that identifies miRNA oligonucleotides, and the results are analyzed to indicate the relative level of miRNA oligonucleotides (and hairpins) that are present in the total RNA of the tissue (Fig. 27).

[0505] Reference is now made to Fig. 26B, which describes how

the microarray to identify miRNA oligonucleotides is designed. miRNA oligonucleotide sequences or potential predicted miRNA oligonucleotides are generated by using known or predicted hairpins as input. Overlapping potential miRNA oligonucleotides are combined to form one larger sub-sequence within a hairpin.

[0506] To generate non-expressed sequences (tails), artificial sequences are generated that are 40 nts in length, which do not appear in the respective organism genome, do not have greater than 40% homology to sequences that appear in the genome, and with no 15-nucleotide window that has greater than 80% homology to sequences that appear in the genome.

[0507] To generate probe sequences, the most probable miRNA oligonucleotide sequences are placed at position 3 (from the 5' end) of the probe. Then, a tail sub-sequence to the miRNA oligonucleotide sequence was attached such that the combined sequence length will meet the required probe length (60 nts for Agilent microarrays).

[0508] The tails method provides better specificity compared to the triplet method. In the triplet method, it cannot be ascertained that the design sequence, and not an uncontrolled window from the triplet probe sequence, was re-

sponsible for hybridizing to the probe. Further the tails method allows the use of different lengths for the potential predicted miRNA oligonucleotide (of combined, overlapping miRNA oligonucleotides).

[0509] Hundreds of control probes were examined in order to ensure the specificity of the microarray. Negative controls contain probes which should have low intensity signal. For other control groups, the concentration of certain specific groups of interest in the library are monitored. Negative controls include tail sequences and non-hairpin sequences. Other controls include mRNA for coding genes, tRNA, and snoRNA.

[0510] For each probe that represents known or predicted miRNA oligonucleotides, additional mismatch probes were assigned in order to verify that the probe intensity is due to perfect match (or as close as possible to a perfect match) binding between the target miRNA oligonucleotide cRNA and its respective complementary sequence on the probe. Mismatches are generated by changing nucleotides in different positions on the probe with their respective complementary nucleotides (A <> T, G <> C, and vice versa). Mismatches in the tail region should not generate a significant change in the intensity of the probe signal, while

mismatches in the miRNA oligonucleotide sequences should induce a drastic decrease in the probe intensity signal. Mismatches at various positions within the miRNA oligonucleotide sequence enable us to detect whether the binding of the probe is a result of perfect match or, alternatively, nearly perfect match binding.

[0511] Based on the above scheme, we designed a DNA microarray prepared by Agilent using their SurePrint technology. Table 11 is a detailed list of microarray chip probes

[0512] KNOWN miRNA OLIGONUCLEOTIDES:

[0513] The miRNA oligonucleotides and their respective precursor sequences are taken from Sanger Database to yield a total of 186 distinct miRNA oligonucleotide and precursor pairs. The following different probes are constructed:

[0514] 1. SINGLE miRNA OLIGONUCLEOTIDE PROBES:

[0515] From each precursor, 26-mer containing the miRNA oligonucleotide were taken, then assigned 3 probes for each extended miRNA oligonucleotide sequence: 1. the 26-mer are at the 5' of the 60-mer probe, 2. the 26-mer are at the 3' of the 60-mer probe, 3. the 26-mer are in the middle of the 60-mer probe. Two different 34-mer subsequences from the design tails are attached to the



26-mer to accomplish 60-mer probe. For a subset of 32 of Single miRNA oligonucleotide probes, six additional mismatches mutations probes were designed:

- [0516] 4 block mismatches at 5' end of the miRNA oligonucleotide;
- [0517] 6 block mismatches at 3' end of the miRNA oligonucleotide;
- [0518] 1 mismatch at position 10 of the miRNA oligonucleotide;
- [0519] 2 mismatches at positions 8 and 17 of the miRNA oligonucleotide;
- [0520] 3 mismatches at positions 6, 12 and 18 of the miRNA oligonucleotide; and
- [0521] 6 mismatches at different positions out of the miRNA oligonucleotide.

[0522] 2. DUPLEX miRNA OLIGONUCLEOTIDE PROBES:

- [0523] From each precursor, a 30-mer containing the miRNA oligonucleotide was taken, then duplicated to obtain 60-mer probe. For a subset of 32 of probes, three additional mismatch mutation probes were designed:
- [0524] 2 mismatches on the first miRNA oligonucleotide;
  - [0525] 2 mismatches on the second miRNA oligonucleotide; and
  - [0526] 2 mismatches on each of the miRNA oligonucleotides.

[0527] 3. TRIPLET miRNA OLIGONUCLEOTIDE PROBES:

[0528] Following Krichevsky's work (Krichevsky et al., RNA 9:1274–1281 (2003)), head to tail ~22–mer length miRNA oligonucleotide sequences were attached to obtain 60–mer probes containing up to three repeats of the same miRNA oligonucleotide sequence. For a subset of 32 probes, three additional mismatch mutation probes were designed:

[0529] 2 mismatches on the first miRNA oligonucleotide;

[0530] 2 mismatches on the second miRNA oligonucleotide; and

[0531] 2 mismatches on each of the miRNA oligonucleotides.

[0532] 4. PRECURSOR WITH miRNA OLIGONUCLEOTIDE PROBES:

For each precursor, 60–mer containing the miRNA oligonucleotide were taken.

[0533] 5. PRECURSOR WITHOUT miRNA OLIGONUCLEOTIDE PROBES:

[0534] For each precursor, a 60–mer containing no more than 16–mer of the miRNA oligonucleotide was taken. For a subset of 32 probes, additional mismatch probes containing four mismatches were designed.

[0535] CONTROL GROUPS:

- [0536] 1. 100 60-mer sequences from representative ribosomal RNAs.
- [0537] 2. 85 60-mer sequences from representatives tRNAs.
- [0538] 3. 19 60-mer sequences from representative snoRNA.
- [0539] 4. 294 random 26-mer sequences from human genome not contained in published or predicted precursor sequences, placing them at the probe's 5' and attached 34-mer tail described above.
- [0540] 5. Negative Control: 182 different 60-mer probes contained different combinations of 10 nt-long sequences, in which each 10 nt-long sequence is very rare in the human genome, and the 60-mer combination is extremely rare.
- [0541] PREDICTED GAM RNAs:
- [0542] There are 8642 pairs of predicted GAM RNA and their respective precursors. From each precursor, a 26-mer containing the GAM RNA was placed at the 5' of the 60-mer probe and a 34-mer tail was attached to it. For each predicted probe, a mutation probes with 2 mismatches at positions 10 and 15 of the GAM RNA were added.
- [0543] For a subset of 661 predicted precursors, up to 2 probes each containing one side of the precursor including any possible GAM RNA in it were added.

[0544] Microarray analysis:

[0545] Based on known miRNA oligonucleotide probes, a preferred position of the miRNA oligonucleotide on the probe was evaluated, and hybridization conditions adjusted and the amount of cRNA to optimize microarray sensitivity and specificity ascertained. Negative controls are used to calculate background signal mean and standard deviation. Different probes of the same miRNA oligonucleotide are used to calculate signal standard deviation as a function of the signal.

[0546] For each probe,  $BG\_Z\_Score = (\log(\text{probe signal}) - \text{mean of } \log(\text{negative control signal})) / (\log(\text{negative control signal}) \text{ standard deviation})$  were calculated.

[0547] For a probe with a reference probe with 2 mismatches on the miRNA oligonucleotide,  $MM\_Z\_Score$   $MM\_Z\_Score = (\log(\text{perfect match signal}) - \log(\text{reference mismatch signal})) / (\text{standard deviation of } \log(\text{signals}) \text{ as the reference mismatch } \log(\text{signal}))$  were calculated.

[0548]  $BG\_Z\_Score$  and  $MM\_Z\_Score$  are used to decide whether the probe is on and its reliability.

[0549] Reference is now made to Fig. 26C, which is a flowchart describing how the cDNA library was prepared from RNA and amplified. The general procedure was performed as

described previously (Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* 2001 15:188-200) with several modifications which will be described hereinbelow.

[0550] First, the starting material is prepared. Instead of starting with standard total RNA, the total RNA was size-fractionated using an YM-100 Microcon column (Millipore Corporation, Billerica, Massachusetts, USA) in the present protocol. Further, the present protocol uses human tissue or cell lines instead of a *Drosophila* in-vitro system as starting materials. Finally, 3 g of size-fractionated total RNA was used for the ligation of adaptor sequences.

[0551] Libraries used for microarray hybridization are listed hereinbelow: "A" library is composed of a mix of libraries from Total HeLa YM100 RNA and Nuclear HeLa YM100 RNA; "B" library is composed of a mix of libraries from Total HEK293 YM100 RNA and Nuclear HEK293 YM100 RNA; "C" library is composed of a mix of YM100 RNA libraries from Total PC3, Nuclear PC3 and from PC3 cells in which Dicer expression was transiently silenced by Dicer specific siRNA; "D" library is prepared from YM100 RNA from Total Human Brain (Ambion Cat#7962); "E" library is prepared from YM100 RNA from Total Human Liver

(Ambion Cat#7960); "F" library is prepared from YM100 RNA from Total Human Thymus (Ambion Cat#7964); "G" library is prepared from YM100 RNA from Total Human Testis (Ambion Cat#7972); and "H" library is prepared from YM100 RNA from Total Human Placenta (Ambion Cat#7950).

[0552] Library letters appended by a numeral "1" or "2" are digested by Xba1 (NEB); Library letters affixed by a numeral "3" are digested by Xba1 and Spe1 (NEB); Library letters appended by a numeral "4" are digested by Xba1 and the transcribed cRNA is then size-fractionated by YM30, retaining the upper fraction consisting of 60 nts and longer; Library letters affixed by a numeral "5" are digested by Xba1 and the transcribed cRNA is then size-fractionated by YM30 retaining the flow-through fraction consequently concentrated with YM10 consisting of 30 nts-60 nts; Library letters affixed by a numeral "6" are digested by Xba1 and the DNA is fractionated on a 13% native acrylamide gel from 40-60 nt, electroeluted on a GeBaFlex Maxi column (GeBa Israel), and lyophilized; Library letters affixed by a numeral "7" are digested by Xba1 and the DNA is fractionated on a 13% native acrylamide gel from 80-160 nt, electroeluted and lyophilized.

[0553] Next, unique RNA-DNA hybrid adaptor sequences with a T7 promoter were designed. This step is also different than other protocols that create libraries for microarrays. Most protocols use complements to the polyA tails of mRNA with a T7 promoter to amplify only mRNA. However, in the present invention, adaptors are used to amplify all of the RNA within the size-fractionated starting material. The adaptor sequences are ligated to the size-fractionated RNA as described in Fig. 22, with subsequent gel-fractionation steps. The RNA is then converted to first strand cDNA using reverse transcription.

[0554] Next, the cDNA is amplified using PCR with adaptor-specific primers. At this point, there is the optional step of removing the tRNA, which is likely to be present because of its low molecular weight, but may add background noise in the present experiments. All tRNA contain the sequence ACC at their 3' end, and the adaptor contains GGT at its 5' end. This sequence together (GGTACC) is the target site for NcoI restriction digestion. Thus, adding the restriction enzyme NcoI either before or during PCR amplification will effectively prevent the exponential amplification of the cDNA sequences that are complements of the tRNAs.

[0555] The amplified DNA is restriction enzyme–digested with Xba1 (and, optionally, with Pst or SpeI) to remove the majority of the adaptor sequences that were initially added to the RNA. Using the first set of RNA–DNA hybrid adaptors listed below, the first two sets of primers listed below, and Xba1 restriction digest yields the following cRNA products: 5'GGCCA – pallindrome/microRNA– UAUCUAG. Using the second set of RNA–DNA hybrid adaptors listed below, the second set of primers listed below, and Xba1 and Pst restriction digest yields the following, smaller cRNA products: 5'GG–pallindrome/microRNA – C\*.

[0556] Then, cDNA is transcribed to cRNA utilizing an RNA polymerase e.g. T7 dictated by the promoter incorporated in the adaptor. cRNA may be labeled in the course of transcription with aminoallyl or fluorescent nucleotides such as Cy3– or Cy5–UTP and CTP among other labels, and cRNA sequences thus transcribed and labeled are hybridized with the microarray.

[0557] The following RNA–DNA hybrid adaptors are included in the present invention:

[0558] Name: T7 NcoI–RNA–DNA 3'Adapter

[0559] Sequence:

5'(5phos)rUrGrGCCTATAGTGAGTCGTATTA(3InvdT)3'



[0560] 2. Name: 5Ada RNA–DNA XbaBseRI

[0561] Sequence: 5' AAAGGAGGAGCTCTAGrArUrA 3' or optionally:

[0562] 3. Name: 5Ada MC RNA–DNA PstAtaBser

[0563] Sequence: 5' CCTAGGAGGAGGACGTCTGrCrArG 3'

[0564] 4. Name: 3'Ada nT7 MC RNA–DNA

[0565] Sequence: 5' (5phos) rCrCrUATAGTGAGTCGTATTATCT (3InvdT)3'

[0566] The following DNA primers are included in the present invention:

[0567] 1. Name: T7 NcoI–RT–PCR primer

[0568] Sequence: 5' TAATACGACTCACTATAGGCCA 3'

[0569] 2. Name: T7NheI SpeI–RT–PCR primer

[0570] Sequence: 5' GCTAGCACTAGTTAATACGACTCACTATAG–GCCA 3'

[0571] 3. Name: 5Ada XbaBseRI Fwd

[0572] Sequence: 5' AAAGGAGGAGCTCTAGATA 3'

[0573] 4. Name: Pst–5Ada XbaBseRI Fwd

[0574] Sequence: 5' TGACCTGCAGAAAGGAGGAGCTCTAGATA 3'

[0575] or optionally:

[0576] 5. Name: 5Ada MC PstAtaBser fwd

[0577] Sequence: 5' ATCCTAGGAGGAGGACGTCTGCAG 3'

[0578] 6. Name: RT nT7 MC Xbal

[0579] Sequence: 5' GCTCTAGGATAATACGACTCACTATAGG 3'

[0580] Reference is now made to Fig. 27A, which demonstrates the detection of known miRNA oligonucleotides and of novel GAM oligonucleotides, using a microarray constructed and operative in accordance with a preferred embodiment of the present invention. Based on negative control probe intensity signals, we evaluated the background, non-specific, logarithmic intensity distribution, and extracted its mean, designated BG\_mean, and standard deviation, designated BG\_std. In order to normalize intensity signals between different microarray experiments, a Z score, which is a statistical measure that quantifies the distance (measured in standard deviations) that a data point is from the mean of a data set, was calculated for each probe with respect to the negative control using the following Z score formula:  $Z = (\text{logarithm of probe signal} - \text{BG\_mean}) / \text{BG\_std}$ . We performed microarray exper-

iments using RNA extracted from several different tissues and we calculated each probes maximum Z score. Fig. 27A shows the percentages of known, predicted and negative control groups that have a higher max Z score than a specified threshold as a function of max Z score threshold. The negative control group plot, included as a reference, considers probe with a max Z score greater then 4 as a reliable probe with meaningful signals. The sensitivity of our method was demonstrated by the detection of almost 80% of the known published miRNA oligonucleotides in at least one of the examined tissues. At a threshold of 4 for the max Z score, 28% of the predicted GAMs are present in at least one of the examined tissues.

[0581] Reference is now made to Fig. 27B, which is a line graph showing specificity of hybridization of a microarray constructed and operative in accordance with a preferred embodiment of the present invention and described hereinabove with reference to Figs. 26A–26C.

[0582] The average signal of known miRNA oligonucleotides in Library A2 is presented on a logarithmic scale as a function of the following probe types under two different hybridization conditions: 50C and 60C: perfect match (PM), six mismatches on the tail (TAIL MM), one mismatch on

the miRNA oligonucleotide (1MM), two separate mismatches on the miRNA oligonucleotide (2MM), three separate mismatches on the miRNA oligonucleotide (3MM). The relative equality of perfect match probes and probes with the same miRNA oligonucleotide but many mismatches over the tail attest to the independence between the tail and the probe signal. At a hybridization temperature of 60°C, one mismatch in the middle of the miRNA oligonucleotide is enough to dramatically reduce the probe signal. Conducting chip hybridization at 60°C ensures that a probe has a very high specificity.

[0583] It is appreciated that these results demonstrate the specificity of the microarray of the present invention in detecting expression of microRNA oligonucleotides.

[0584] Reference is now made to Fig. 27C, which is a summary table demonstrating detection of known microRNA oligonucleotides using a microarray constructed and operative in accordance with a preferred embodiment of the present invention and described hereinabove with reference to Figs. 26A–26C.

[0585] Labeled cRNA from HeLa cells and Human Liver, Brain, Thymus, Placenta, and Testes was used for 6 different hybridizations. The table contains the quantitative values

obtained for each miRNA oligonucleotide probe. For each miRNA oligonucleotide, the highest value (or values) is given in bolded font while lower values are given in regular font size. Results for MIR-124A, MIR-9 and MIR-122A are exactly as expected from previous studies. The References column contains the relevant references in the published literature for each case. In addition to these miRNA oligonucleotides, the table shows other known miRNA oligonucleotides that are expressed in a tissue-specific manner. We show that MIR-128A, MIR-129 and MIR-128B are highly enriched in Brain; MIR-194, MIR-148 and MIR-192 are highly enriched in Liver; miR-96, MIR-150, MIR-205, MIR-182 and MIR-183 are highly enriched in Thymus; MIR-204, MIR-10B, MIR-154 and MIR134 are highly enriched in Testes; and MIR-122, MIR-210, MIR-221, MIR-141, MIR-23A, MIR-200C and MIR-136 are highly enriched in Placenta. In most cases, low but significant levels are observed in the other tissues. However, in some cases, miRNA oligonucleotides are also expressed at relative high levels in an additional tissue.

[0586] It is appreciated that these results reproduce previously published studies of expression of known microRNA oligonucleotides. These results demonstrate the reliability

of the microarray of the present invention in detecting expression of published microRNA oligonucleotides, and of novel GAM oligonucleotides of the present invention.

## **DETAILED DESCRIPTION OF TABLES**

[0587] Table 1 comprises data relating the SEQ ID NO of oligonucleotides of the present invention to their corresponding GAM NAME, and contains the following fields: GAM SEQ-ID: GAM SEQ ID NO, as in the Sequence Listing; GAM NAME: Rosetta Genomics Ltd. nomenclature (see below); GAM RNA SEQUENCE: Sequence (5' to 3') of the mature, "diced" GAM RNA; GAM ORGANISM: identity of the organism encoding the GAM oligonucleotide; GAM POS: Dicer cut location (see below); and

[0588] Table 2 comprises detailed textual description according to the description of Fig. 8 of each of a plurality of novel GAM oligonucleotides of the present invention, and contains the following fields: GAM NAME: Rosetta Genomics Ltd. nomenclature (see below); GAM ORGANISM: identity of the organism encoding the GAM oligonucleotide; PRECUR SEQ-ID: GAM precursor Seq-ID, as in the Sequence Listing; PRECURSOR SEQUENCE: Sequence (5' to 3') of the GAM precursor ; GAM DESCRIPTION: Detailed description of GAM oligonucleotide with reference to Fig. 8; and

[0589] Table 3 comprises data relating to the source and location of novel GAM oligonucleotides of the present invention, and contains the following fields: GAM NAME: Rosetta Genomics Ltd. nomenclature (see below); PRECUR SEQ-ID: GAM precursor SEQ ID NO, as in the Sequence Listing; GAM ORGANISM: identity of the organism encodes the GAM oligonucleotide; SOURCE: Chromosome encoding a human GAM oligonucleotide ; STRAND: Orientation of the strand, "+" for the plus strand, "-" for the minus strand; SRC-START OFFSET: Start offset of GAM precursor sequence relative to the SOURCE; SRC-END OFFSET: End offset of GAM precursor sequence relative to the SOURCE; and

[0590] Table 4 comprises data relating to GAM precursors of novel GAM oligonucleotides of the present invention, and contains the following fields: GAM NAME: Rosetta Genomics Ltd. nomenclature (see below); PRECUR SEQ-ID: GAM precursor Seq-ID, as in the Sequence Listing; GAM ORGANISM: identity of the organism encoding the GAM oligonucleotide; PRECURSOR-SEQUENCE: GAM precursor nucleotide sequence (5' to 3'); GAM FOLDED PRECURSOR RNA: Schematic representation of the GAM folded precursor, beginning 5' end (beginning of upper row) to 3' end

(beginning of lower row), where the hairpin loop is positioned at the right part of the draw; and

[0591] Table 5 comprises data relating to GAM oligonucleotides of the present invention, and contains the following fields: GAM NAME: Rosetta Genomics Ltd. nomenclature (see below); GAM ORGANISM: identity of the organism encoding the GAM oligonucleotide; GAM RNA SEQUENCE: Sequence (5' to 3') of the mature, "diced" GAM RNA ; PRECUR SEQ-ID : GAM precursor Seq-ID, as in the Sequence Listing; GAM POS: Dicer cut location (see below); and

[0592] Table 6 comprises data relating SEQ ID NO of the GAM target gene binding site sequence to TARGET gene name and target binding site sequence, and contains the following fields: TARGET BINDING SITE SEQ-ID: Target binding site SEQ ID NO, as in the Sequence Listing; TARGET ORGANISM: identity of organism encode the TARGET gene; TARGET: GAM target gene name; TARGET BINDING SITE SEQUENCE: Nucleotide sequence (5' to 3') of the target binding site; and

[0593] Table 7 comprises data relating to target-genes and binding sites of GAM oligonucleotides of the present invention, and contains the following fields: GAM NAME: Rosetta Genomics Ltd. nomenclature (see below); GAM



ORGANISM: identity of the organism encoding the GAM oligonucleotide; GAM RNA SEQUENCE: Sequence (5' to 3') of the mature, "diced" GAM RNA; TARGET: GAM target gene name; TARGET REF-ID: Target accession number (GenBank); TARGET ORGANISM: identity of organism encode the TARGET gene; UTR: Untranslated region of binding site/s (3' or 5'); TARGET BS-SEQ: Nucleotide sequence (5' to 3') of the target binding site; BINDING SITE-DRAW: Schematic representation of the binding site, upper row represent 5' to 3' sequence of the GAM, Lower row represent 3' to 5' Sequence of the target; GAM POS: Dicer cut location (see below); and

[0594] Table 8 comprises data relating to functions and utilities of novel GAM oligonucleotides of the present invention, and contains the following fields: GAM NAME: Rosetta Genomics Ltd. nomenclature (see below); GAM RNA SEQUENCE: Sequence (5' to 3') of the mature, "diced" GAM RNA; GAM ORGANISM: identity of the organism encoding the GAM oligonucleotide; TARGET: GAM target gene name; TARGET ORGANISM: identity of organism encode the TARGET gene; GAM FUNCTION: Description of the GAM functions and utilities; GAM POS: Dicer cut location (see below); and

[0595] Table 9 comprises references of GAMs target genes and contains the following fields: TARGET: Target gene name; TARGET ORGANISM: identity of organism encode the TARGET gene; REFERENCES: reference relating to the target gene; and

[0596] Table 10 comprises data relating to novel GR (Genomic Record) polynucleotides of the present invention, and contains the following fields: GR NAME: Rosetta Genomics Ltd.nomenclature (see below); GR ORGANISM: identity of the organism encoding the GR polynucleotide; GR DESCRIPTION: Detailed description of a GR polynucleotide, with reference to Fig. 16; and

[0597] Table 11 comprises data of all sequences printed on the chip experiment as described herein above with reference to Fig 26 and include the following fields: PROBE SEQUENCE: the sequence that was printed on the chip PROBE TYPE: as described in details in Fig. 26 in chip design section and summarized as following: a. Known – published miR, Known\_mis1 – published miR with 1 mismatch mutation on miR sequence. Known\_mis2 – published miR with 2 mismatches mutation on miR sequenced. Known\_mis3 – published miR with 3 mismatches mutation on miR sequence, Known\_mis4 – published miR with 6 mismatches

mutation not on miR sequence, Predicted – GAM–Rosetta Genomics Ltd. Mismatch – GAM–Rosetta Genomics Ltd. with 2 mismatches, Edges1 – left half of GAM–Rosetta Genomics Ltd, Edges2 – right half of GAM–Rosetta Genomics Ltd extended with its palindrom, Control1 – negative control, Control2 – random sequences, l. Control3 – tRNA, m. Control4 – snoRNA, Control5 – mRNA, Control6 – other.;GAM RNA SEQ ID/MIR NAME: for GAM–Rosetta Genomics Ltd. Nomenclature(see below); GAM RNA SEQUENCE: Sequence (5' to 3') of the mature, "diced" GAM RNA; LIBRARY: the library name as defined in Fig. 26C; SIGNAL: Raw signal data for library ; BACKGROUND Z-SCORE: Z score of probe signal with respect to background, negative control signals; MISMATCH Z-SCORE: Z-score of probe signal with respect to its mismatch probe signal;

[0598] Table 12 comprises data relating to diseases that GAM oligonucleotides are predicted to regulate the disease-associated genes. Each row is referred to a specific disease, and lists the GAM target genes related to the disease. The first row is a summary of ALL diseases containing in the present invention, thus listing ALL GAM target genes relating to theses diseases. The table contains the

following fields: ROW#: index of the row number; DISEASE NAME: name of the disease; TARGET-GENES ASSOCIATED WITH DISEASE: list of GAM target genes that are associated with the specified disease; and

[0599] Table 13 comprises data related to the GAM RNA SEQUENCES included in the present invention that were validated by laboratory means. If the validated sequence appeared in more than one GAM precursor, the GAM RNA SEQ-ID indicated may be arbitrarily chosen. The VALIDATION METHOD indicates the type of validation performed on the sequence: "Mir Sequencing" refers to miRNA oligonucleotide sequences that were sequenced, as described hereinabove with reference to Fig. 22. Other validations are from microarray experiments as described hereinabove with reference to Figs. 26A-C and 27A-C. The SIGNAL indicates a raw signal data; BACKGROUND Z-SCORE indicates a Z score of probe signal with respect to background, negative control signals; MISMATCH Z-SCORE: indicates a Z-score of probe signal with respect to its mismatch probe signal. The microrray validations are divided into two groups: a) "Chip strong" refers to miRNA oligonucleotide sequences whose intensity (SIGNAL) on the microarray "chip" was more than 6 standard deviations

above the background intensity, and the differential to the corresponding mismatch intensity was more than 2 standard deviations, where in this case the standard deviation is of the intensity of identical probes and b) "Chip" refers to miRNA oligonucleotide sequences, whose intensity was more than 4 standard deviations above the background intensity.

[0600] Table 14 comprises sequence data of GAMs associated with different diseases. Each row refers to a specific disease, and lists the SEQ ID NOs of GAMs that target genes associated with that disease. The table contains the following fields: ROW#: index of the row number; DISEASE NAME: name of the disease; SEQ ID NOs OF GAMs ASSOCIATED WITH DISEASE: list of sequence listing IDs of GAMs targeting genes that are associated with the specified disease; and

[0601] The following conventions and abbreviations are used in the tables: The nucleotide "U" is represented as "T" in the tables, and;

[0602] GAM NAME or GR NAME are names for nucleotide sequences of the present invention given by RosettaGenomics Ltd. nomenclature method. All GAMs/GRs are designated by GAMx/GRx where x is a unique ID.

[0603] GAM POS is a position of the GAM RNA on the GAM PRE-CURSOR RNA sequence. This position is the Dicer cut location: A indicates a probable Dicer cut location; B indicates an alternative Dicer cut location.

[0604] All human nucleotide sequences of the present invention as well as their chromosomal location and strand orientation are derived from sequence records of UCSC-hg16 version, which is based on NCBI, Build34 database (April, 2003).